EFFECTS OF TASK TRAINING ON KINDERGARTEN STUDENTS' PERFORMANCE ON EARLY LITERACY MEASURES

by

Sara Moore Mackiewicz

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Special Education

Charlotte

2010

Approved by:

____
Dr. Nancy Cooke

____
Dr. Charles Wood

____
Dr. Claudia Flowers

____
Dr. Michael Thompson

ABSTRACT

SARA MOORE MACKIEWICZ. Effects of task training on kindergarten students' performance on early literacy skills. (Under the direction of DR. NANCY L. COOKE)

The use of early literacy screening measures helps determine which students are at risk for future reading difficulties. However, there has been some recent concern related to the classification validity of screening measures (Hintze, Ryan, & Stoner, 2003; Nelson, 2008). Low classification validity results in the identification of a large number of false positives, students who are falsely identified as being at risk. Task training may help to address false positive rates by providing brief instruction focused on helping students understand demands and expectations of the measure.

This true experimental study investigated the effects of task training for three DIBELS subtests (i.e., Initial Sound Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency) in order to differentiate the need for supplemental instruction from task misunderstanding for students in kindergarten. Participants were randomly assigned to either the treatment group or the control group and change in instructional status recommendation between pretest and posttest were examined along with the change in score on the individual subtests. Results indicated that students in the treatment group (n=20) were significantly more likely to move up in instructional status. On the pretest, all students in both groups demonstrated the need for supplemental instruction. Based on results of the posttest, only 35% of the treatment group still demonstrated the need for supplemental instruction while 82% of students in the control group still demonstrated the need for extra support. Additionally, students in the treatment group outperformed the control group (n=22) when a combination of subtest performance was examined.

DEDICATION

I would like to dedicate this work to several members of my family who provided love and support throughout this process. First, to my father, James Riley Moore: you instilled in me a love of reading and learning and never allowed me to doubt my abilities. Although you are not present to witness this accomplishment, I can hear your laughter and feel your pride. To my mother, Pat Moore: you have always provided unconditional love and support and you showed me that it is never too late to realize your dreams.

Finally, to my husband and my son: Brad, there are no words to express the depth of my gratitude for all you have done. Thank you for your unwavering dedication, support, and selflessness. Over the past five years, many people have asked me how I manage to do what I do. The only answer I have is you. I could not have done it without you. To Jack: Your smile, laughter, and hugs inspired me throughout this journey. I love you more than words can say.

ACKNOWLEDGEMENTS

I would like to express my appreciation to several people who have been vital to my success during my time at UNC Charlotte. First, I would like to thank my committee members for their time, patience, expertise, and input. I would like to express sincere gratitude to my advisor, Nancy Cooke, for her dedication and investment in my success. She pushed me to realize my potential as a teacher, a researcher, and a leader and serves as a role model as I begin my career in higher education. I am in awe of her dedication to students and to the field of special education. I would also like to thank Charlie Wood for providing countless opportunities to grow professionally, offering practical advice every step of the way, and answering countless questions.

Additionally, I am thankful to Claudia Flowers for her invaluable input related to design and statistics and to Michael Thompson for taking the time to engage in a project unrelated to his discipline. I also want to express sincere appreciation to my friend and colleague, Tara Galloway. We began this journey together and I appreciate her encouragement, support, and friendship throughout the past five years. Finally, I am thankful for the kindergarten students who made this project worthwhile and so much fun!

TABLE OF CONTENTS

CHAPTER1: INTRODUCTION 1

CHAPTER 2: REVIEW OF THE LITERATURE 18

CHAPTER 3: METHOD 57

Participants 57

Setting 58

Researcher 59

Interventionists and Assessors 60

Independent Observer 61

Instrumentation 62

Interscorer Reliability 66

Dependent Variables 67

Procedures and Data Collection 67

Social Validity 79

Research Design 81

Data Analysis 81

CHAPTER 4: RESULTS 83

Treatment Fidelity 83

Assessment Integrity and Interobserver Reliability 85

Effects of Task Training on Recommendation Status 86

Effects of Task Training on Individual Subtest Performance 89

Social Validity Results 93

CHAPTER 5: DISCUSSION 97

Effects of Task Training on Dependent Variables 98

Discussion of Social Validity Data 104

Limitations 105

Recommendations for Future Research 107

Conclusion 109

REFERENCES 110

APPENDIX A: TASK TRAINING SCRIPTS 123

APPENDIX B: TAST TRAINING FIDELITY CHECKLISTS 130

APPENDIX C: SOCIAL VALIDITY QUESTIONNAIRE 136

LIST OF TABLES

TABLE 1: Demographic information for participants	58
TABLE 2: Order of task training administration	72
TABLE 3: Timeline for data collection	80
TABLE 4: Mean percentage of interscorer reliability across subtests	86
TABLE 5: Percentages at each risk level across measures	87
TABLE 6: Instructional status recommendation change by group	89
TABLE 7: Descriptive statistics for each measure by group	90
TABLE 8: ANOVA results of posttest scores by group	92

CHAPTER 1: INTRODUCTION

Statement of Problem

The National Assessment of Educational Progress (NAEP) describes the academic achievement of our nation's students and indicates that 37% of fourth graders in the United States cannot read on grade level. Even more alarming is the finding that 70% of low income students in the fourth grade are unable to read at a basic level (National Center for Education Statistics [NCES], 2006). Unfortunately, reading difficulties begin much earlier than fourth grade.

In fact, kindergarten students enter school with "meaningful differences" in vocabulary knowledge (Hart & Risley, 1995). For example, Hart and Risley found that a preschool child in a middle-class family hears approximately 11 million words per year, while a child in a low-income family hears approximately 3 million words per year. By age 4, the gap in words heard grows to 13 million for a child in a low-income family versus 45 million for a child in a middle-class family. Hart and Risley continued their research to investigate whether vocabulary knowledge at age 3 predicted language skills when the children reached third grade. Results indicated that vocabulary use at age 3 was strongly associated with scores on measures of receptive vocabulary, overall language development, and reading comprehension when measured in third grade.

Stanovich (1986) applied the Matthew Effect to describe how children continue to fall behind their peers in regard to reading achievement. He indicated that, the "rich get richer" and the "poor get poorer," meaning that students experiencing early reading difficulties continue to fall behind their peers as they progress through their school-age years.

The Matthew Effect is supported by data from a longitudinal study conducted by Juel (1988) that found if a child was a poor reader at the end of first grade there was an 88% chance that the child would still be a poor reader at the end of fourth grade. Furthermore, that child is also likely to experience continued difficulty with reading when in ninth grade (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Shaywitz, Escobar, Shaywitz, Fletcher, and Makuch, 1992). Additional research shows that students not meeting grade level expectations by third grade are likely to never catch up to their peers (Farkas, 2003; Manset-Williamson, St. John, Hu, & Gordon, 2002). In many cases, appropriate interventions may not be implemented until age nine and approximately 75% of students will continue to experience reading difficulties through their high school years (Lyon, 1998), continue to fail, and may demonstrate a need for special education services sometime during their school-age years (Simmons, Kame'enui, Coyne, & Chard, 2002). This research demonstrates that it is imperative for school personnel to address any skill deficits early in a student's school career. Early intervention is needed in order to address the critical needs of students who are at risk for developing reading difficulties.

Current federal and state legislation reflects the urgency of prevention and intervention by mandating that all students make progress. In response, initiatives have been developed that strive to address low academic achievement. One of these initiatives, Response to Intervention (RTI), includes the use of differentiated instruction, usually through the implementation of a three-tiered model. In an RTI framework, intervention is provided at increasing levels of intensity, or tiers, and a continuum of instructional services are employed in order to address a student’s specific skill deficits (Brown-Chidsey & Steege, 2005; Burns & VanDerHeyden, 2006; Griffiths, VanDerHeyden, Parson, & Burns, 2006).

One of the key features of RTI is the use of a universal screening measure to identify students at risk for reading failure; these students may need supplemental instruction provided at a higher tier or intensity level. Many students experience difficulty in acquiring basic reading skills during their early school years. Fortunately, with early intervention most reading problems can be prevented (Snow, Burns, & Griffin, 1998). However, in order to make sound instructional decisions, accurate assessment of the most important early literacy skills is necessary (Coyne & Harn, 2006). The most important early literacy skills are those that predict future reading achievement. Therefore, early assessments should especially focus on these skills.

Several efforts, including the work from the National Research Council (Snow et al., 1998) and Adams (1990) have been undertaken to determine the most important beginning reading skills. These reviews of research confirmed that reading achievement is impacted by a student’s proficiency with skills that build the foundation for later success in reading including (a) phonological awareness, (b) alphabetic understanding, (c) accuracy and fluency with connected text, (d) vocabulary, and (e) comprehension. These five foundational skills have been referred to as the “big ideas” in beginning reading (Simmons et al., 2002).

A large body of research has investigated possible predictors of future reading success. Several skills related to the “big ideas” of beginning reading are known, strong predictors of this success including phonemic awareness, letter naming skills, and alphabetic understanding (Good, Simmons, & Kame'enui, 2001; O'Connor & Jenkins, 1999; Torgesen, Wagner, Rashotte, Rose, Lindamood, Conway et al., 1999). Phonemic awareness is the ability to hear and manipulate the sounds in spoken words and the understanding that spoken words and syllables are made up of sequences of speech sounds (Yopp, 1992). Phonemic awareness is one of the most accurate predictors of future reading achievement. Phonemic awareness plays a causal role in the acquisition of beginning reading. Research has shown that the primary difference between good and poor readers is the good reader's superior phonological processing ability (Adams, Foorman, Lundberg, & Beeler, 1998; NRP, 2000; Pepper & Felton, 1995).

Another important early literacy indicator is alphabetic understanding or alphabetic principle. A student has acquired an understanding of the alphabetic principle when he or she demonstrates the ability to associate sounds with written letters (Moats, 1999; Torgesen, 2002). When a student uses these associations to blend sounds and read words, the student is decoding. Decoding is a necessary strategy for reading the English language because there are too many words to simply memorize them all (Bay Area Reading Task Force, 1996) and to become a proficient reader a student must have a strategy to decode, or read, words (NRP, 2000).

In order to make informed instructional decisions and to ensure that students are acquiring the necessary prerequisite literacy skills, educators must have appropriate measurement tools available for screening students, monitoring student progress toward early literacy goals, and evaluating the effectiveness of instructional programs. Unfortunately, for the past two decades, there has been growing concern and dissatisfaction with static assessments that measure student knowledge at one point in time. During the administration of traditional assessments the examiner is considered an objective observer and does not actively intervene during testing (Caffrey, Fuchs, & Fuchs, 2008; Haywood & Tzuriel, 2002) and, in addition, these approaches to assessment provide limited feedback and/or practice and offer no scaffolding for learning how to complete the task (Campione, 1989; Embertson, 1992; Fuchs, Fuchs, Compton, Bouton, Caffrey, & Hill, 2007).

In addition, traditional assessment procedures may not identify students who simply need more assistance with understanding the directions of the task. Explanations for this confusion have been proposed by several researchers in the field including a misunderstanding of directions (Campbell & Carlson, 1995; Haywood, Brown, & Wingenfeld, 1990) and linguistic and cultural bias (Lopez, 1997). It may be difficult to differentiate between students who truly need extra support to learn and students who simply did not comprehend the task they were being asked to respond to. With a universal screening administered to all students on a standardized, norm-referenced assessment, false positives may be identified. A false positive occurs when a student who eventually becomes a proficient reader scores below the cut score on the screening instrument and is falsely identified as at risk for academic failure (Fuchs et al., 2007). These falsely identified students mean more school resources are consumed in order to provide intervention to students who may not need it (Fletcher, Foorman, Boudousquie, Barnes, Schatschneider, & Francis, 2002).

Dynamic assessment (DA) may be able to help solve some of the problems associated with traditional assessments. DA or “learning potential assessment,” can provide mediated learning that is responsive to a student’s specific, identified needs

(Moore-Brown, Huerta, Uranga-Hernandez, & Pena, 2006). This concept is the ability to benefit from a learning experience which leads to a change in performance on similar tasks. These methods include a group of approaches that are linked by the common component of building instruction and feedback into the assessment process. With these approaches, instruction and feedback are differentiated on the basis of an individual's performance on the assessment (Elliott, 2003).

Procedures and outcome goals vary among the different DA methods. Several of the most common methods include Feuerstein's Learning Potential Assessment Device (LPAD; Feuerstein, Rand, & Rynders, 1988), Budoff's Learning Potential Testing (Budoff, 1967, 1987a, 1987b; Budoff & Friedman, 1964), graduated prompts (Campione & Brown, 1987; Campione, Brown, Ferrara, Jones, & Steinberg, 1985), the information-processing framework (Swanson, 1995), and testing-the-limits procedures (Carlson & Wiedl, 1978, 1979). Each of these methods and the research investigating their effectiveness are discussed in greater detail within the literature review of Chapter 2.

Overall, DA measures more than the performance of skills at one point in time because it allows the examiner to administer a pretest, provide instruction related to the student's performance and then administer a posttest (Lidz & Pena, 1996; Olswang & Bain, 1996). DA procedures attempt to change performance by offering individualized scaffolded assistance in an effort to understand a student's true learning potential (Swanson & Lussier, 2001) and can help account for variables that may underestimate an individual's ability (e.g., unfamiliarity with the task, language, or materials; Haney & Evans, 1999). Few research studies have investigated the use of DA as an independent variable, but what has been done appears promising. Results from these studies (Cormier, Carlson, & Das, 1990; Fagundes, Haynes, Haak, & Moran, 1998; Kalyuga & Sweller, 2005; Kar, Dash, Das, & Carlson, 1993; Missiuna & Samuels, 1989; Tzuriel & Shamir, 2002) indicate that DA procedures may be valuable in helping to address the growing concern regarding use of standardized, static assessments, including insensitivity to what the learner can do given some support (e.g., prompting, drawing attention to important features of the problem or task).

Precorrection is another strategy for preventing excessive errors and improving future student performance through abbreviated instruction. Instead of designing the intervention based on test performance, as with DA, precorrection anticipates potential problems and addresses them prior to instruction. For example, in using precorrections to prevent challenging behavior, Colvin, Sugai, Good, and Lee (1997) followed a series of seven steps including (a) determining a potential obstacle, (b) making clear the expected behavior, (c) changing the context, (d) modeling the expected behavior, (e) reinforcing the expected behavior, (f) providing motivation to perform the expected behavior, and (g) monitoring performance. This intervention aims to prevent challenging behavior (errors) rather than reacting to the challenging behaviors after they occur.

The majority of research investigating the effectiveness of precorrection strategies has been in the area of behavior management (e.g., Colvin et al., 1997; Lewis, Colvin, & Sugai, 2000; Oswald, Safran, & Johanson, 2005), but this strategy may also have the potential to increase academic achievement. Recently, there has been some research investigating the effects of precorrection on reading outcomes including examination of a decoding precorrection strategy (Miao, Darch, & Rabren, 2002) and a case study involving the preview of key words and reading passages (O'Donnell, Weber, &

McLaughlin, 2003). In each case, probable errors were prevented by providing brief instruction before students engaged in the full task. Although precorrections share some attributes with DA (i.e., prevention of errors with brief instructional support), no literature has been identified in which precorrections were applied to the assessment of academic tasks. A second difference is that DA is conducted individually in response to previous errors, whereas precorrections may be used with a group to prevent typical errors.

In order to determine if DA has the potential to reduce the number of identified false positives, the effectiveness of the test-teach-retest format (DA procedures) when used as an intervention needs to be evaluated. The studies investigating the use of DA in the area of cognitive ability have shown DA as effective when used by specific practitioners. The procedures and types of measures used in these studies are feasible for use by psychologists and/or diagnosticians, but are generally not practical for use by teachers, due to the individualized, sustained interaction with a single student. In addition, DA procedures may not be appropriate for screening purposes because they are individualized to each student's specific needs as identified with a pretest and conducted over time, which prevents generalizability of treatment protocols to typical screening conditions in which many students must be assessed.

Precorrection procedures also may help in the identification of false positives by helping students avoid making common errors on assessment measures. Precorrection implemented by teachers has been investigated in previous studies and this strategy appears practical for classroom use, but may not be appropriate for use prior to screening administration. Precorrections target specific items that will be encountered during an instructional activity rather than familiarizing students with task expectations.

However, task training, an approach that combines features of both DA and precorrection may help to address some of the difficulties associated with using either of the interventions in conjunction with a universal screening tool. Task training can be described as an abbreviated combination of dynamic assessment and precorrection that provides efficient instruction (e.g., explicit, brief) focused on helping students understand the task demands. Task training may reduce the number of false positives identified by a universal screening when errors are due to lack of clarity regarding the task. Insufficient improvement following task training may indicate a true need for remediation, and as a result, task training has the potential to lead to more informed, and thereby more accurate, instructional decision making for students at risk. Research is needed in order to determine if task training can help make this differentiation between students in need of supplemental instruction and those who are not.

One study has investigated the effects of task training procedures to differentiate the need for supplemental instruction from task misunderstanding on an early literacy assessment tool (Mackiewicz, Cooke, Galloway, & Helf, 2010). A randomized pretest-posttest experimental design was used to compare the effects of task training on the phoneme segmentation skills of kindergarten students. The treatment group received brief task training, while the control group received no intervention other than instruction within the general education classroom. Significant differences were found between the two groups on a posttest measure of phoneme segmentation fluency.

Purpose of the Study

In the Mackiewicz et al. (2010) study, only one early literacy skill, phoneme segmentation, was targeted during that investigation. Most schools use a combination of four measures at mid-year kindergarten to determine the need for supplemental instruction. The purpose of the current research was to empirically investigate the effects of task training targeting four early literacy measures in order to differentiate the need for supplemental instruction from task misunderstanding for kindergarten students.

Significance of the Study

In order to make informed instructional decisions, educators must have appropriate measurement tools available for screening students, monitoring student progress toward early literacy goals, and evaluating the effectiveness of instructional programs. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Kaminski & Good, 1996, 1998), a measurement system developed by researchers at the University of Oregon, is widely used to evaluate students' early literacy skills and monitor progress toward benchmark goals. Appropriate levels of reliability and validity for screening, monitoring progress, and evaluating the outcomes of instructional programs have been established for the DIBELS and will be discussed in greater detail in Chapter 2 (Good, Gruba, & Kaminski, 2001).

Administration of the DIBELS subtest measures is standardized, and is, therefore susceptible to the same problems exhibited by static, traditional assessment measures. Accurate assessment of students' early literacy skills, especially skills that are predictive of future reading achievement, is necessary in order to make appropriate instructional decisions and provide supplemental instruction that matches students' needs. Assessment tools serve many purposes, one of which is to facilitate appropriate instruction. This study will identify an intervention that can increase the accuracy of assessment results used for identifying kindergarten students who need supplementary reading instruction.

The current investigation is important because it may lead to identification of an efficient and effective task training protocol that could be used to ensure correct placement of students, reducing the number of students identified as needing supplemental instruction. Task training may lead to decreased numbers of false positives, or students who, according to assessment results, seem to need supplemental instruction, but in reality scored poorly because they did not understand the task demands (Fuchs et al., 2007). Reducing the number of falsely identified students may lead to more appropriate allocation of financial resources for schools, including school personnel and materials. Determining the effects of task training is important because it may be a low cost intervention (brief, limited personnel, limited materials) that has the potential to offer substantial benefit to schools. In addition, task training my also benefit students. If students are placed in the appropriate instructional groups, then they may make larger academic gains because instruction will be focused on the skills they need to learn and not skills they have already aquired.

Research Questions

This study empirically investigated the effects of task training with kindergarten students, targeting four early literacy measures, in order to differentiate the need for supplemental instruction from task misunderstanding. Specifically, this study addressed the following research questions with kindergarten students whose combination of pretest scores on the DIBELS mid-year benchmark measures resulted in recommendations for supplementary instruction (i.e., strategic or intensive support):

1. Will there be a significant difference between students who receive task training and students who do not receive training on the DIBELS instructional status recommendation for supplemental instruction?
2. Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of initial sounds isolated on the DIBELS Initial Sound Fluency subtest?
3. Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of phonemes identified on the DIBELS Phoneme Segmentation Fluency subtest?
4. Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of letter sounds identified on the DIBELS Nonsense Word Fluency subtest?
5. What are kindergarten teachers' perceptions regarding the acceptability of the task training procedures?

Definitions

Terms that were used in the study and their definitions are presented in the following section. The terms that were chosen will be critical for understanding the implementation procedures and observed results.

Alphabetic awareness: Knowledge of letters of the alphabet along with the understanding that the alphabet represents the sounds of spoken language and the correspondence of spoken sounds to written language (http://reading.uoregon.edu).

Alphabetic principle: the ability to associate sounds or phonemes with letters and use these sounds to read words (Torgesen, 2002).

Alphabetic understanding: Understanding that the left-to-right spellings of printed words represent their phonemes from first to last (http://reading.uoregon.edu).

At-risk for reading failure: A term used to identify students who perform below grade level on basic reading skills (Foorman & Torgesen, 2001).

Automaticity: The ability to translate letters-to-sounds-to-words fluently, effortlessly (Snow et al., 1998).

Blending: A process by which students listen to a sequence of separately spoken phonemes and then combine the phonemes to form a word (Center for the Improvement of Early Reading Achievement, 2001).

Core reading program: The primary instructional tool that teachers use to teach children to learn to read and ensure they reach reading levels that meet or exceed grade-level standard (http://reading.uoregon.edu).

Curriculum-based measurement (CBM): A method of assessment that can be used to determine how students are progressing in basic academic areas such as math, reading, writing, and spelling. CBM describes a students' academic competence at a single point in time, quantifies the rate at which the student develops academic competence over time, and provides information designed to help educators design more effective programs to increase student achievement (Deno, Fuchs, Marston, & Shin, 2001).

Decoding: The process of figuring out an unfamiliar word by breaking it into individual sounds. Readers use knowledge about letter-sound relationships and the alphabetic code to decode words (http://reading.uoregon.edu).

Dynamic assessment: An interactive approach to conducting assessments within the domains of psychology, speech/language, or education, that focuses on the ability of the learner to respond to intervention (Elliott, 2003; Moore-Brown et al., 2006).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A set of procedures and measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade. They are designed to be short fluency measures used to regularly monitor the development of early literacy and early reading skills (Good & Kaminski, 2002).

False positive: Occurs when a student who eventually becomes a proficient reader scores below the cut score on a screening instrument and is falsely identified as at risk for academic failure (Fuchs et al., 2007).

Letter-sound correspondence: The link between a letter or combination of letters and a sound (Coyne, Kame'enui, & Simmons, 2001).

Most common sound: The sound a letter most frequently makes in a short, one syllable word (e.g., the sound of g in rag; the sound of e in red; http://reading.uoregon.edu).

Nonsense word: A word in which the letters make their most common sounds but the word has no commonly recognized meaning (e.g., lat, ut; (http://reading.uoregon.edu).

Phoneme: The smallest unit of sound that can be combined to form syllables and words (Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001).

Phonemic awareness: the ability to hear and manipulate the sounds in spoken words and the understanding that spoken words and syllables are made up of sequences of speech sounds (Yopp, 1992).

Phonics: the understanding that there is a predictable relationship between phonemes (the sounds of spoken language) and graphemes (the letters and spellings that represent those words in written language (http://reading.uoregon.edu).

Phonological awareness: The ability to hear and manipulate the sound structure of language. This is an encompassing term that involves working with the sounds of language at the word, syllable, and phoneme level (Ehri et al., 2001).

Precorrection: An antecedent instructional event designed to prevent the occurrence of predictable problem behavior and to facilitate the occurrence of more appropriate replacement behavior (Colvin et al., 1997).

Response to intervention (RTI): Combines assessment and intervention within a multi-level prevention system to maximize student achievement. Using RTI, schools identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions, and adjust the intensity and nature of those interventions depending on a student's responsiveness (National Center on Response to Intervention, 2009).

Screening: An inventory that provides the teacher a beginning indication of a student's preparation for grade level reading instruction. It is a "first alert" that a child may need extra help to make adequate progress in reading (http://www.fcrr.org).

Segmentation: A process by which students break a word into its separate sounds, saying each sound aloud (Center for the Improvement of Early Reading Achievement, 2001).

Supplemental instruction: Instruction that goes beyond that provided by the comprehensive core program because the core program does not provide enough instruction or practice in a key area and provides additional coverage and extra practice of the necessary components of reading (http://www.fcrr.org).

Task training: Abbreviated combination of dynamic assessment and precorrection that provides efficient instruction (e.g., explicit, brief) focused on helping students understand the task demands (Mackiewicz et al., 2010).

Delimitations

This study will be delimited by geographical restrictions to an urban school in a southeastern state. In addition, the participants will be selected for inclusion in the study because they will have scored below identified benchmarks on the DIBELS measures administered at the mid-year of kindergarten. Therefore, generalizations can only be made to DIBELS kindergarten mid-year benchmark measures of ISF, PSF, and NWF. Also, generalizations can only be made to kindergarten students who are identified through a combination of these same mid-year benchmark subtests as needing strategic or intensive support.

Summary

In summary, accurate assessment of students' early literacy skills, including phonemic awareness, letter naming fluency, and alphabetic principle, is necessary in order to make appropriate instructional decisions. This study will empirically investigate the use of task training on early literacy indicators to differentiate the need for supplemental instruction from task misunderstanding for kindergarten students. In addition, this study will add to the limited research regarding the effectiveness of task training procedures.

Chapter 2 will provide a review of the related literature important to this study and Chapter 3 will provide a description of the methodology that will be used. Chapter 4 will provide a summary of the results and a discussion of implications will be provided in Chapter 5.

CHAPTER 2: REVIEW OF LITERATURE

Early Intervention

Previously, there was a widespread belief that students acquired reading skills through a natural progression of development and that young students would be ready to read at different points in time. This belief was reported by Fletcher, Satz, and Morris (1984) and Satz, Taylor, Friel, and Fletcher (1978). As a result, slow acquisition of beginning reading skills was considered natural for some students and intervention for struggling readers was traditionally not offered until third or fourth grade. However, as noted in Chapter 1, students who do not learn to read proficiently at a young age often do not catch up to their peers (Francis et al., 1996; Juel, 1988; Stanovich, 1986). Research conducted over the last three decades has shown that early intervention is critical and we now know that successful early literacy experiences are important for later reading success. One way to prevent reading failure is to implement effective interventions with young children, before they fall too far behind their classmates (Snow et al., 1998). This section is a review of that literature.

This research has shown that students falling behind early in their school careers are less likely to catch up to their peers (Francis et al., 1996; Juel, 1998). Francis et al. demonstrated the importance of learning to read in early grades through a longitudinal study addressing reading skill development from kindergarten through ninth grade. Results indicated that the majority of students who were poor readers in third grade did not “catch up” to their peers. In fact, 74% of students who were poor readers in third grade continued to be poor readers in ninth grade.

Another longitudinal study conducted by Juel (1998) followed 54 children from first grade through fourth grade and found that students who are poor readers at the end of first grade face an 88% probability of continuing to be poor readers at the end of fourth grade. In addition, assessment results indicated that children identified as poor readers at the end of first grade entered first grade with little phonemic awareness.

Many studies have investigated the effectiveness of early reading intervention. Most of these studies have focused on phonemic awareness and phonics skills with students in kindergarten and first grade. This research has shown positive effects for students in kindergarten (Blachman, Ball, Black, & Tangel, 1994; O’Connor, Jenkins, & Slocum, 1995; O’Connor, Notari-Syverson, & Vadasy, 1996) and first grade (Blachman, Tangel, Ball, Black, & McGraw, 1999; Clay, 1985; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Vellutino et al., 1996).

Wanzek and Vaughn (2007) conducted a review of the literature focused on studies published between the years 1995 and 2005 investigating early reading intervention for students with reading difficulties and disabilities. The review included 18 studies and focused on interventions occurring in kindergarten through third grade that were implemented for at least 100 sessions. The authors targeted several aspects of the studies including student outcomes following participation in an extensive early reading intervention and intervention features associated with high effect size (e.g., duration, level of standardization).

Results of the review indicated overall positive reading outcomes for students with reading difficulties participating in the various extensive interventions investigated in the nine studies. Specifically, effect sizes were greater in studies where a smaller intervention group size was used and when intervention was provided early, mostly in kindergarten and first grade.

Since 2005, several additional studies have been published that focus on early intervention. Vellutino, Scanlon, Small, and Fanuele (2006) investigated the importance of kindergarten intervention. At the beginning of kindergarten, 30% of the students assessed were identified as at risk for future reading difficulty using a screening tool. Approximately half of the at-risk students were randomly assigned to the treatment group and the other half were assigned to a school-based comparison group. In kindergarten, participants received small-group (i.e., 2 to 3 students) early literacy intervention from a certified teacher twice a week for 30-min sessions. The intervention consisted of many activities focused on important early literacy skills such as print awareness, letter recognition, letter identification, phonological awareness, letter-sound correspondence, and sight word reading. At the end of kindergarten, all students were assessed on several phonologically-based literacy skills including phoneme segmentation, letter names, letter sounds, word identification, spelling, and letter-sound decoding. A statistically significant difference was found between the treatment and comparison group, with the treatment group outperforming the comparison group. These results indicate that early intervention for students identified as at-risk for reading failure can significantly improve their early reading skills when they are identified at the beginning of kindergarten.

Students who participated in the kindergarten study described above were reevaluated at the beginning of first grade using six early literacy measures. A composite score was used to differentiate between students who remained at risk for reading difficulties and students who were no longer at risk. Identified students were randomly assigned to one of three conditions, two treatment groups and one comparison group. Both treatment groups included instruction in both phonological skills and text processing skills, but each group spent the majority of time on only one of these areas. That is, one group was provided one-to-one instruction focused on the development of phonological skills, while a second group received one-to-one instruction focused on development of text processing skills. The third group received intervention that was typically available at their home schools. Intervention ended at the conclusion of first grade, but participants were assessed at the end of first, second and third grades.

Eighty-four percent of students who received intervention, in kindergarten only or received intervention in kindergarten and first grade, performed within the average range on all literacy measures administered at the end of third grade. Furthermore, 73% of students in the treatment groups performing within the average range received intervention only in kindergarten. The authors concluded that long-term reading difficulties may be preventable for the majority of students identified as at-risk at the beginning of kindergarten. Some participants still needed intervention at the beginning of first grade, but most of those students were no longer in need of supplemental instruction by the end of first grade.

Another study that began with intervention in kindergarten was conducted by O’Connor, Fulmer, Harty, and Bell (2005). A longitudinal-lagged design was used in two schools to investigate the effects of two layers of intervention with treatment groups compared to an historical control group. Layer one consisted of professional development for school personnel focused on research related to early literacy instruction, activities for teaching each reading component, and ongoing support provided by research staff. Layer two included direct intervention for students identified as being at risk for reading difficulties. In year one, kindergarten students identified as at risk were provided small-group instruction (i.e., 2 to 3 students) for 10 to 15 min, 3 times per week. In year two, first graders who met the criteria for at-risk status were provided with small group instruction for approximately 25 min, 3 times per week. Direct intervention was continued during years three and four for students who continued to be identified as at risk. Results indicated improved reading outcomes for students who participated in early and sustained intervention. For students who began receiving intervention in year one of the study, the authors found a decrease in the incidence of identified reading disability by the end of third grade, when compared to the historical control group.

The research discussed above supports the notion that early intervention is needed before a child experiences too much failure and they are unable to catch up to their peers. Another important reason for early intervention is the difficulty associated with remediation for students experiencing reading failure in the later grades. Several aspects make later remediation more difficult. One of the most important being that remediation requires more time and resources than intervention provided during the early learning years. In fact, research sponsored by the National Institute of Child Health and Development (NICHD; 2000) indicates that it takes as much as four times the assistance to improve a student's reading ability if help is provided in fourth grade compared to intervention that begins in the middle of kindergarten (Hall & Moats, 1999). This "wait to fail" model of identification required students to fail for many years before remediation was provided and this led to an increased number of students in need of special education (Simmons et al., 2002) and an overrepresentation of minority students in special education (Chinn & Hughes, 1987; Hosp & Reschly, 2003).

This combined evidence documents the need for current federal legislation that mandates student progress and increases accountability for school districts, schools, and individual teachers. First, the *No Child Left Behind Act* (NCLB) was signed into law in 2001. Part of this law stressed the need for early intervention, for the prevention of reading difficulties, assessment of student outcomes, increased accountability, and professional development that assists teachers with implementation of evidence-based instructional practices. The 2004 reauthorization of the *Individuals with Disabilities Education Act* (IDEA, 2004) supports the mandates of NCLB in several ways, including the use of scientifically-based reading strategies and prevention of academic difficulties through early intervention. This legislation represents a shift from intervention or remedial models to a more proactive approach by providing instruction for students at risk for developing serious reading difficulties.

In summary, there is extensive research supporting early prevention and intervention of reading difficulties and reading disabilities. These studies indicate that early instructional intervention can make a difference in reading outcomes for students. Furthermore, research shows that interventions implemented with kindergarten and first grade students identified as at risk appear to be the most effective in preventing reading failure. The culmination of decades of research in this area led to changes in federal legislation, which has led the educational community to focus on proactive models of intervention.

Response to Intervention

Response to Intervention (RTI) is a multi-step approach to providing intensive intervention to at-risk learners for the prevention of future academic difficulties and can also be used as a way to identify students with learning disabilities. RTI is a promising approach for addressing the needs of all students who are exhibiting learning difficulties in the general education setting by helping to accomplish several goals including (a) early identification of students at risk for academic difficulties through universal screening practices, (b) early and targeted intervention for students identified at risk, (c) progress monitoring practices to assist with data-based instructional decision making, (d) use of increasingly more intensive tiers of evidence-based instruction, and (e) increased confidence that students referred for special education services are not struggling due to inadequate or inappropriate instruction (Fletcher, Coulter, Reschly, & Vaughn, 2004).

Although various models of RTI implementation exist, in order to accomplish the above stated goals there are three core concepts that are common across all models. These core concepts include: (a) systematic application of scientific, research-based intervention in the general education setting; (b) systematic measurement of a student's response to these interventions; and (c) the use of data to inform instructional decisions (Brown-Chidsey & Steege, 2005; Burns & VanDerHeyden, 2006; Griffiths et al., 2006).

Furthermore, the National Research Center on Learning Disabilities (NRCLD; 2007) has identified several components necessary for strong and effective implementation of RTI. Two of these components are related to assessment. First, implementation of a school-wide, universal screening is necessary in order to determine which students are at risk for the development of future learning problems. Another necessary feature is the use of research-based progress monitoring. This is a set of assessment procedures that aids in determining whether or not a student is making progress while in a specific intervention program. Progress monitoring allows for data driven instructional decision making.

In an RTI approach, schools provide intervention at increasing levels of intensity. Typically there are three levels, referred to as tiers, that offer a continuum of instructional strategies and services to students exhibiting skill deficits (Burns & VanDerHeyden, 2006; National Joint Committee on Learning Disabilities, 2005; Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007). Tier 1 is the core instructional program that is characterized by high quality instruction for all students in the general education setting and is intended to meet the instructional needs of most students. At this level, all students are screened to identify which students may need additional intervention. Students whose rate of progress is behind that of their peers move to Tier 2. At this level, the goal is to meet the needs of students identified as at risk by providing supplemental instruction in small groups in an effort to support and reinforce the skills being taught within the core reading program. Progress monitoring occurs frequently at this level to further assist in instructional decision making. A small percentage of students fail to respond sufficiently to Tier 2 intervention, so they move to Tier 3. This tier provides instruction that is more explicit, intensive, and individualized.

As RTI is a relatively new concept, researchers are just beginning to develop frameworks and conduct in depth investigations into the effectiveness of various RTI models. In 2007, Hollenbeck published a review of the literature related to the implementation of RTI models. Thirty-six studies were reviewed and 29 of those studies were published in 2003 or after. As a result of the review, several potential benefits to RTI implementation emerged, including emphasized general education accountability, fewer students identified as being disabled, and increased collaboration and shared responsibility across general and special education. Benefits substantiated by research included the benefit of early intervention for struggling readers (Vellutino et al., 2006) and the benefit of reducing teacher bias in referral (Speece & Case, 2001), subsequently leading to the reduction of a disproportionate number of minority students identified as disabled.

Several recent studies illustrate the usefulness of RTI as an early identification and intervention model. These studies have examined the implementation effects on special education referrals and the number of students identified as eligible for special education services. A study conducted by O'Connor, Harty, and Fulmer (2005) examined the effects of interventions at Tiers 2 and 3 as students progressed from kindergarten through third grade. Researchers focused on participants' reading development and special education placement rates by third grade. Historical control groups of third graders attending the two target schools were used during the investigation. Tier 1 consisted of ongoing professional development for teachers in addition to the core reading program. Small-group instruction provided 3 days per week was provided as Tier 2 intervention and began after the administration of a screening measure at midyear. Tier 3 consisted of individual or small group (i.e., 2 students) instruction provided 5 days per week.

Results of this investigation compared the reading achievement of third grade students who were identified as at risk in kindergarten to the historical control group. Moderate to large effect sizes (ES = 0.4 to 1.8) were found in the areas of decoding, word identification, fluency, and reading comprehension, favoring children in the treatment group who participated in the tiered interventions. In addition, the incidence of special education placement in the historical control group averaged 15%, while, following participation in the tiered instruction for 4 years, the treatment group placement rate was 8%.

Another study investigating the implementation of an RTI model over time (through use of two cohorts) was conducted by Vaughn, Wanzek, Murray, Scammacca, Linan-Thompson, and Woodruff (2009). This study specifically focused on students who made limited progress in both Tiers 1 and 2, or *low responders*. The RTI model used in this study consisted of three instructional tiers. Tier 1, or primary instruction, included the core reading curriculum paired with professional development provided to teachers and screening of all students three times per year. Tier 2, or secondary intervention, included implementation of supplemental instruction provided daily for 30 min and progress monitoring for students identified as at risk. Tier 3, or tertiary intervention, included intensive intervention for students demonstrating a “low response” to the second intervention phase or Tier 2.

Prior to intervention, all first grade students were screened and those identified as at risk were randomly assigned to treatment and comparison groups as part of a larger longitudinal study being conducted by the research team. Students in the treatment group received intervention from the research team, while students in the comparison group participated in the typical school reading programs. At the end of first grade, students in the treatment group who met pre-existing criteria were considered *high responders* and intervention was not provided for those students in second grade. However, students identified as *low responders* following Tier 2 intervention were provided intensive instruction at Tier 3 (i.e., 50 min daily in groups of 3 students) for all of second grade.

The effectiveness of the tertiary intervention in this study was examined using a regression-discontinuity research design in order to compare the performance of the *high responders* to the performance of the *low* responders. Following intervention in second grade, the students were assessed on several measures of reading. The researchers examined the performance of *low responders* at the end of second grade relative to the performance of *high responders* that participated in the same Tier 2 intervention in first grade, but did not participate in Tier 3 intervention in second grade. In the areas of reading comprehension and word reading, significant differences were found between the two groups with the *low responders* outperforming the *high responders.* No significant differences were found between the two groups in the area of reading fluency.

A different approach to investigating the effectiveness of an RTI model was conducted by VanDerHeyden, Witt, and Gilbertson (2007). This study not only examined the effect of the model on student outcomes but also examined effects on: (a) the number of evaluations for special education eligibility conducted and the percentage of students who qualified for services; (b) the degree that the data generated from the RTI process influenced the decisions made by a school team making special education referrals; (c) the identification rates by ethnicity, gender, socio-economic status, and primary language; and (d) the assessment and placement costs for the district. A multiple-baseline across schools design was used to investigate implementation of a specific RTI model identified as STEEP or System to Enhance Educational Performance. STEEP is described as a systematic, research-based model of RTI that uses a series of assessment and intervention procedures with specific decision rules in place to identify students who might benefit from an evaluation to determine eligibility for special education services. For the purpose of this study, the STEEP model was implemented in five elementary schools serving grades 1 through 5 over a 2 year period.

Following the first two years of implementation the data were analyzed in an effort to answer the research questions. Fewer evaluations for the purpose of determining special education eligibility were conducted and the students evaluated were more likely to qualify for special education services when STEEP data were included in the referral team's decision making. After one year of STEEP implementation, students identified as in need of special education due to a Specific Learning Disability (SLD) fell from 6% to 3.5% for elementary-age students district-wide. During baseline the percentage of minority students ranged from 2 to 5% for all participating schools and these percentages were maintained (approximately 3%) during STEEP implementation. When gender was examined, a disproportionate number of males were evaluated and determined eligible for special education during baseline. After STEEP was implemented, the total number of evaluations was reduced with a more pronounced reduction for male students positively affecting disproportionality. Finally, a cost analysis indicated that resources devoted to traditional assessment were reduced and replaced by direct assessment, intervention, and consultation services.

In summary, RTI further demonstrates the effectiveness of early identification of students at risk for future reading difficulties and/or disabilities and the subsequent early intervention to address identified deficits. Based on tentative evidence, several potential benefits exist when RTI models are implemented within elementary school settings. First, students participating in a tiered model of instruction appear to outperform students in comparison or control groups on various measures of reading outcomes. Second, the number of evaluations conducted to determine eligibility for special education and the number of students deemed eligible for these services seems to be reduced within an RTI model of instruction. Third, RTI implementation appears to positively affect the disproportionate placement of male students in special education programs. All of these benefits culminate in an important consideration for districts and individual schools, especially in light of the current economic situation in our country. RTI implementation may reduce the time spent on unnecessary evaluations, therefore reducing district costs.

Early Literacy Assessment

The first step in determining which students are in need of early reading intervention is through early literacy skill assessment. Due to the importance of assessment and early identification highlighted in the above discussions related to early intervention and RTI, valid and reliable assessment tools are imperative. The literacy skills acquired in kindergarten are the foundation for the development of future reading ability, so we must be able to accurately assess these skills in order to improve instruction and intervention (Coyne et al., 2001). Because of this, identification of students in need of additional intervention should start at the beginning of kindergarten and continue throughout elementary school using measures that target indicators of future reading achievement (Good, Simmons, & Smith, 1998; Snow et al., 1998). The use of screening assessments helps determine which students are at risk for future reading difficulties and in need of intervention.

Although we know its importance, screening in order to identify students at risk for future reading difficulties is not always straightforward and simple. Speece (2005) wrote that the challenge for the early identification of students at risk for reading disabilities is finding a screening tool that can hit a “moving target known as reading development.” Many times screening leads to implementation of intensive and expensive instructional practices in an effort to move students out of the at risk category (e.g., Torgesen, Alexander, Wagner, Rashotte, Voelher, & Conway, 2001; Vellutino et al., 1996), so it is necessary that the assessment tools used, accurately identify students in need of that intervention.

In order to assist with the challenges of identifying appropriate screening tools and other assessment measures, the Reading First Assessment Committee (RFAC) was formed. The goal of this committee was to provide state and local education agencies with guidance in the selection and use of reading assessment tools. In order to accomplish this goal, the RFAC developed criteria for evaluating the adequacy of reading measures used in kindergarten through third grade and compared widely used assessment instruments to the committee’s developed criteria (Kame’enui et al., 2006).

The committee described screening as “brief assessments conducted with all children, typically at the beginning of the school year. It targets skills that are strongly predictive of important future reading outcomes” (Kame’enui et al., 2006; p. 4). The committee’s report also indicated the need for screening measures to attend to sensitivity and specificity. Additional essential features of accurate screening tools are predictive power and usefulness in making instructional decisions (Good et al., 2001).

Predictive power is the ability of assessment instruments to accurately and reliably identify those students most likely to experience future reading difficulties. A large amount of research has been done investigating possible predictors of future reading success. Several skills related to the big ideas of beginning reading have been determined as strong predictors of future reading achievement including phonemic awareness, letter naming skills, and alphabetic understanding. The following is a review of some of that literature.

Early literacy predictors. Phonemes are the smallest distinguishable unit of spoken language (Mathes & Torgesen, 1998) and phonemic awareness is the ability to hear and manipulate the individual sounds in spoken words. Research related to investigating the connection between phonemic awareness and reading acquisition has shown this to be an important relationship (Ball & Blachman, 1991; Ehri et al., 2001; Stanovich, 1986). In fact, phonemic awareness has been identified as one of the most accurate predictors of future reading achievement.

Specifically, experimental and longitudinal studies have revealed that phonemic awareness and letter knowledge are the strongest predictors of reading skill acquisition (Wagner, Torgesen, & Raschotte, 1994). In 1986, Stanovich published a literature review focused on this topic and reported that phonemic awareness is a more powerful predictor of future reading achievement than nonverbal intelligence, vocabulary, and listening comprehension. In addition, a literature review focused on reading research (Smith, Simmons, & Kame’enui, 1998) found convincing evidence that phonological awareness plays an important role in beginning reading acquisition. The most distinguishing characteristic of children with learning disabilities in reading when compared to peers without disabilities appears to be phonological processing deficits (Wagner et al., 1997; Wolf & Bowers, 1999).

In addition to phonemic awareness, predictive validity studies also indicate that assessment measures focused on letter-naming speed are especially effective for identifying students at risk for future reading difficulties (O'Connor & Jenkins, 1999; Wagner et al., 1997). This skill is one of the best predictors of future reading achievement (Bond & Dykstra, 1967; Share, Jorm, Maclean, & Matthews, 1984). Letter-naming fluency has also been shown to predict later word reading ability (Daly, Wright, Kelly, & Martens, 1997; Kaminski & Good, 1996; Walsh, Price, & Gillingham, 1988).

Research has shown that phonemic awareness is critical for young students, but this skill alone is not enough (Gough & Tunmer, 1986). Another important early literacy indicator is alphabetic principle or alphabetic understanding (Schatschneider & Torgesen, 2004; Snow et al., 1998). Alphabetic principle establishes a clear link between a letter and a sound and requires a reader to understand that the letters of our alphabet (i.e., graphemes) are directly connected to sounds (i.e., phonemes). A student has acquired an understanding of alphabetic principle when they demonstrate the ability to associate sounds with written letters (Moats, 1999; Torgesen, 2002). When a student uses these associations to blend sounds and read words, the student is decoding.

More recently, Schatschneider, Fletcher, Francis, Carlson, and Foorman (2004) used a cross-sequential longitudinal design to follow several cohorts of students from the beginning of their kindergarten year to either the end of first grade or the end of second grade. The purpose of their study was to identify important predictors of early reading performance (in kindergarten) and examine how the identified predictors related to subsequent reading achievement. At the beginning of kindergarten, students were evaluated on the following skills (a) letter names, (b) letter sounds, (c) phonological awareness, (d) oral language skills, and (e) rapid automatized naming of letters and objects. A total of 10 predictors were measured at the beginning of kindergarten and 8 predictors were measured during three additional assessment periods of that school year. Outcome assessments administered at the end of first and second grades measured word recognition, reading comprehension, and reading fluency. Following data analysis, the researchers concluded that measures of phonological awareness, letter-sound knowledge, and naming speed consistently accounted for the unique variance across reading outcomes in both grades 1 and 2.

Numerous studies investigating the predictive validity of various combinations of these three important early literacy skills (i.e., phonemic awareness, rapid letter naming, alphabetic principle) have been conducted. Blachman (1984) investigated the relationship of phonological awareness skills, specifically segmenting and rhyming, and rapid naming abilities to future reading achievement in kindergarten and first grade. Rapid naming abilities and phonological awareness abilities were related to at least half of the outcome measures used in kindergarten. When evaluated at the end of first grade, rapid naming abilities and phonological awareness skills were predictive of the reading outcome scores.

In addition, results indicated that students with deficiencies in rapid naming abilities did not necessarily demonstrate a deficit in the area of phonological awareness.

Approximately a decade later, MacDonald and Cornwall (1995) published results from a study that investigated the relationship between phonological awareness skills in kindergarten and reading and spelling achievement 11 years later. Results indicated that phonological awareness was a long-term predictor of both word identification skills and spelling ability. In addition, phonological awareness in kindergarten was a better predictor of reading comprehension at age 17 years than was word identification and spelling achievement in kindergarten.

As a result of consensus reports (e.g., Snow et al., 1998) and previous research discussed above we know that the assessment of these predictor skills is important and should begin in kindergarten. However, even with all the evidence supporting the assessment of phonological awareness skills, alphabetic principle, and rapid naming skills, the reading research community is still searching for the early reading assessments that are most predictive of future reading ability (Bishop, 2003). Researchers are beginning to conduct predictive validity studies in order to determine *the* most important early literacy skills to assess.

One example of such studies was conducted by Stage, Sheppard, Davidson, and Browning (2001). These researchers investigated the predictiveness of kindergarten students' performance on letter-naming and letter-sound fluency measures using growth curve analysis. Outcome reading measures included an oral reading fluency measure administered at various times throughout first grade. Students' first-grade growth in oral reading fluency was significantly predicted by both letter-naming and letter-sound fluency performance.

Another study conducted by Bishop (2003) in an effort to identify a specific combination of predictive measures that correlate with later reading achievement, examined the accuracy of the measures, and determined the optimal time frame in which to administer assessments in kindergarten. Over 100 kindergarten students from three schools participated in the study, which lasted for 2 years. Measures of letter identification, phonological awareness, phonological memory, and rapid automatized naming were administered at the beginning and at midyear during kindergarten. Reading outcomes were measured when students reached the end of first grade. Analysis included five predictive models composed of various combinations of the predictor variables. Results of the analysis indicated that the model combining letter identification, phonological awareness, and rapid automatized naming was the best predictor of early reading achievement. There was no significant difference between the beginning and midyear assessment windows in terms of predictability.

In summary, research has demonstrated specific skills that, when measured in kindergarten, can predict future reading achievement, indicating that measurement tools need to address the assessment of these important skills. Turning attention back to the RFAC’s report (Kame’enui et al., 2006), many assessments were reviewed, but only a few of these met the committee’s developed standards for overall quality and technical adequacy. Even though no measures met all of the evaluation criteria, the committee determined that the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good &

Kaminski, 2002) is an assessment tool with sufficient scientific evidence for use as an early literacy screening measure, progress monitoring tool, and outcome assessment.

Dynamic Indicators of Basic Early Literacy Skills

The DIBELS are a set of procedures and measures used to assess the acquisition of early literacy skills for students in kindergarten through sixth grade. It is a universal screening tool that is widely used to identify children as at risk for reading difficulties (U.S. Department of Education, Office of Inspector General, 2003). They are designed to be short, fluency measures used to regularly monitor the development of important early literacy skills and measure empirically validated skills related to future reading achievement (Good & Kaminski, 2002). The measures were specifically designed to assess the core components of reading including phonological awareness, alphabetic principle, accuracy and fluency with connected text, vocabulary, and comprehension. Additionally, DIBELS measures have demonstrated technical adequacy based on evidence of reliability, validity, and sensitivity to small changes in skills.

The DIBELS measurement system consists of the following subtests: Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Word Use Fluency (WUF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Oral Reading Fluency (ORF), and Retell Fluency (RTF). Administration of the benchmark subtests occur three times per year, at the beginning, midyear, and end of the school year. Various configurations of these subtests are administered according to grade level and time of year. For example, at the beginning kindergarten benchmark only the ISF and LNF subtests are administered. At the midyear benchmark, kindergarten students are assessed using the ISF, LNF, PSF, and NWF subtests. The WUF subtest may also be administered;

however, a benchmark goal for this subtest is not provided by the developers because additional research is needed to establish its linkage to other big ideas of literacy. The end of year kindergarten administration includes the LNF, PSF, and NWF subtests.

Two of the measures administered at midyear kindergarten focus on the assessment of phonemic awareness. ISF assesses a child's skill at identifying and producing the initial sound of a given word and PSF assesses the child's skills at producing the individual sounds within a given word. The LNF subtest measures the rate of letter naming. NWF is a measure of alphabetic principle and assesses a child's knowledge of letter-sound correspondences and their ability to blend letters together to form unfamiliar, "make-believe" words.

Good, Simmons, Kame'enui, Kaminski, and Wallin (2002) developed a technical report describing the decision rules for instructional recommendations and how the rules were developed. The researchers followed general rules and principles when determining the decision rules including (a) establishing cutoffs and goals where the odds would be in favor of reaching future early literacy goals and (b) identifying students who were unlikely to achieve future early literacy goals without intervention. For the individual measures, the researchers identified a level of performance at which the odds were in favor for the student to achieve future literacy skills, a performance level where the students were more likely to not achieve future reading goals, and a middle category where performance within an identified range did not predict future reading achievement either way. If the measure is administered prior to the benchmark goal, performance on that subtest is categorized as ***low risk***, ***some risk***, or ***at risk.*** If the measure is administered at the time of the benchmark goal, the subtest performance is referred to as ***established***, ***emerging***, or ***deficit.***

When all of the subtest scores for a particular administration period have been entered into the DIBELS Data system, an overall instructional recommendation is made. If the overall performance indicates odds in favor of achieving future reading goals a recommendation of ***Benchmark - At grade level*** is made. If the student's performance indicates odds against achieving those future reading goals the instructional recommendation is ***Intensive - Needs Substantial Intervention***. When a student's performance does not provide a clear prediction of future reading achievement the instructional recommendation is ***Strategic - Additional Intervention***.

Several studies have investigated the predictive validity of these specific DIBELS measures (Burke, Hagan-Burke, Kwok, & Parker, 2009; Elliott, Lee, & Tollefson, 2001; Rouse & Fantuzzo, 2006) and a review of those studies is discussed next.

Elliott et al. (2001) examined a modified version of the DIBELS subtests using letter naming fluency, sound naming fluency, initial phoneme ability, and phoneme segmentation ability with 75 kindergarten students. The concurrent criterion-related validity of the modified DIBELS measures was examined and significant correlations were found between predictor variables and criterion achievement measures, ranging from .12 to .81. However, these findings should be interpreted with caution because the modified version of two of the subtests were accuracy based, as opposed to the fluency-based design of the original and current DIBELS subtests.

A more recent investigation conducted by Rouse and Fantuzzo (2006) examined the convergent and predictive validity of three DIBELS subtests administered at the end of kindergarten. The subtests included LNF, PSF, and NWF and scores were analyzed along with standardized scores from outcome reading measures administered at the end of first grade. Results of a canonical correlation analysis indicated significant predictive relationships between the early literacy skills measured at the end of kindergarten and the literacy constructs measured at the end of first grade. More specifically, LNF was strongly associated with the structure of the *Test of Early Reading Ability* (Reid, Hresko, & Hammill, 2001). All three DIBELS subtests taken together explained approximately 52% of the variance in instructional reading from the *Developmental Reading Assessment* (Beaver, 1997). LNF appeared to be the strongest predictor of instructional reading level, followed by NWF and PSF.

Most recently, Burke et al. (2009) used path analysis to investigate which early literacy indicators from kindergarten DIBELS, if any, can be used to model reading acquisition and what the predictor relationships are when the measures are ordered in developmental progression. Researchers developed the proposed model based on a theoretical model of developmental progression of reading acquisition, where skills build on one another eventually resulting in reading fluency and text comprehension (Ehri, 1995).

In this study, the following DIBELS measures were administered to 218 kindergarten students at midyear: ISF, PSF, LNF, and NWF. Outcome measures were also administered across the duration of the study (i.e., 3 years) including measures of phonemic decoding efficiency, sight word efficiency, oral reading fluency, and passage comprehension. Following data analysis, the results of the model fit indicated that performance on all four of the midyear kindergarten DIBELS subtests was valid in predicting the more complex alphabetic skills required for reading achievement. The researchers concluded that the results of their study provided strong support for the predictive validity of the DIBELS subtests.

In addition to the numerous studies investigating the predictive validity of the DIBELS subtests, one recent study examined the intervention validity of the DIBELS PSF measure (Hagans, 2008). In this study, the PSF and NWF subtests were used to monitor the acquisition of literacy skills for 75 first grade students. Students were randomly assigned to either a treatment group receiving early literacy instruction or a control group. The independent variables examined during the investigation included socioeconomic status of student families and instructional program participation. The effects of instructional group on early literacy skills as measured by PSF were examined using hierarchical linear modeling (HLM). The study's findings support the practice of using results from the PSF subtest to inform instructional planning, which subsequently resulted in increased phoneme segmentation skills for participants.

In summary, effective early intervention is driven by the ability to make sound instructional decisions. These decisions can only be made with data obtained through reliable and valid assessment tools. For early literacy development, in particular, there is an abundance of research that has identified several early literacy skills that, when measured as early as kindergarten, can predict future reading achievement. Because of this, predictive power is also an extremely important feature for early literacy assessments to demonstrate. Based on the research reviewed in this section, it appears that the DIBELS have been shown to be reliable, valid, and predictive of future reading ability.

False Positives

In addition to the features discussed above (i.e., reliability, validity, and predictability), assessment tools should also demonstrate diagnostic accuracy. A universal screening tool should identify students truly at risk (i.e., true positives) while limiting the number of students falsely identified as at risk (i.e., false positives). Jenkins (2003) recommended that within models of early identification, including RTI models, assessment tools should yield a high percentage of true positives (e.g., sensitivity rates above 90%). High sensitivity rates allow for the identification of a manageable risk pool by limiting the number of false positives identified.

However, assessment measures used for the identification of students at risk for reading difficulties and/or disabilities in the early grades, including DIBELS, purposefully “overidentify” students as at risk to ensure that all students who possibly have problems in early literacy skill development will be provided with early support (Wanzek & Vaughn, 2007). Overidentifiying students leads to the likelihood of the identification of more false positives with less severe difficulties in kindergarten and first grade samples than in higher grades.

In evaluating reading assessments, most of the evidence examined by the RFAC related to criterion validity derived from correlational data, but classification validity has been identified as possibly being more important information than criterion validity when determining the usefulness of reading screening measures (Bishop, 2003; Jenkins, 2003). Investigations into classification validity compare the number of examinees identified as exhibiting or not exhibiting problems on a “gold standard” test as compared with the number of examinees identified at risk on a screening measure, which can also be described as the identification of true positives and true negatives. The gold standard, or reference standard, is considered to be “the best available evidence for the existence of a particular condition or characteristic” (Kessel & Zimmerman, 1993).

Classification validity had not been investigated and/or reported for most of the assessment tools reviewed by the committee, including the DIBELS. However, since the RFAC report was published in 2001, at least two studies investigating the classification validity of DIBELS have been conducted (Hintze, Ryan, & Stoner, 2003; Nelson, 2008).

The study by Hintze et al. (2003) investigated several research questions. First, they examined the concurrent validity of the DIBELS with the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) when both measures were administered to 86 kindergarten students midyear. Results showed moderate to strong correlations between the DIBELS and the CTOPP suggesting that these assessments measure a similar construct. For the purpose of the current review, the authors’ purpose related to examining the classification validity of the DIBELS will be discussed in greater detail. A decision accuracy study based on suggested cut-scores and cut-scores determined as a result of receiver operator characteristic (ROC) curve analysis was conducted. Results of the analysis indicated that using the test developer’s suggested cut-scores resulted in “extremely high sensitivity” with low levels of specificity. Specifically, for both the ISF and PSF subtests, use of the suggested cut-scores resulted in a very high percentage of true positives, but this came at the expense of a large number of false positives.

Nelson (2008) extended the examination of the classification validity of the DIBELS with kindergarten students. The DIBELS subtests of ISF, LNF, PSF, and NWF were administered along with a norm-referenced test of phonological awareness, the *Test of Phonological Awareness – Second Edition: Plus* (TOPA-2+; Torgesen & Bryant, 2004) to the students at midyear. The same students were administered a second norm-referenced test, this one focused on reading skills (Woodcock-Johnson Tests of Achievement, Third Edition; WJ III; Woodcock, McGrew, & Mather, 2001), at the end of kindergarten.

Receiver operator characteristic (ROC) curve analysis was used to assess the overall diagnostic accuracy of each DIBELS subtest, while the area under the curve (AUC) was calculated in additional analysis. Analyses resulted in sensitivity indexes within the 80 to 90% range for ISF, PSF, and NWF. Sensitivity rates for LNF ranged from 53 to 72%. For the ISF, PSF, and NWF subtests, the false positive rates ranged from 41 to 72%. All AUC indexes for the DIBELS subtests indicated medium overall diagnostic accuracy when the WJ III was used as the reference standard.

Nelson (2008) also examined the diagnostic accuracy of the DIBELS subtests when used together. When examining the cut-scores for the *at risk* level, sensitivity rates were over 85% and moderate specificity rates were found. Use of the *some risk* cut-scores resulted in perfect sensitivity rates (i.e., 100%), but, as in the Hintze et al. study, came at the expense of a very high false positive rate (86 to 88%).

The results of these studies led researchers to investigate why the false positive rates for the DIBELS subtests are so high. One study focused on this issue was conducted by Catts, Petscher, Schatschneider, Bridges, & Mendoza (2009). The purpose of their study was to examine the distributions of scores obtained from the DIBELS after being administered to a large group (>18,000) of children at various points during their early elementary school years. The researchers were also interested in the presence of floor effects and their impact on the predictive validity of the DIBELS. The term "floor effects" refers to artificially restricting how low the scores can be resulting in many students performing near the lower end of the distribution. Floor effect can lead to the over identification of at-risk students. The subtests of ISF, LNF, PSF, NWF, and ORF were included in the investigation and compared to performance on the Reading Comprehension subtest of the 10th edition of the *Stanford Achievement Test* (SAT-10; Harcourt Educational Measurement, 2003).

Results indicated that each DIBELS subtest initially showed strong floor effects, but, over time, these effects lessened. Quantile regression plots showed that there were low to moderate correlations between performance on the DIBELS subtests and the outcome measure (i.e., SAT-10); however, for most of the DIBELS subtests, predictability improved across administrations.

In addition, logistic regression analysis was used to investigate the accuracy of the DIBELS in predicting which children will become good or poor readers as measured by a reading outcome measure. Administration of the ISF and PSF subtests resulted in high false positive rates (>.50). The LNF and NWF subtests appeared to be better predictors of reading outcomes than ISF and PSF. The authors concluded that floor effects affected the ability of these measures to accurately predict future reading outcomes.

Much of this chapter has focused on the importance of early identification for early intervention, but substantial false positive rates can prevent the accurate early identification of students at risk for future reading failure and should be kept to a minimum (Compton, Fuchs, Fuchs, & Bryant, 2006; Compton, Fuchs, Fuchs, Elleman, &

Gilbert, 2008). In order for early intervention models to be effective, screening measures should identify true positives at a rate of at least 90% (Jenkins, 2003). High false positive rates interfere with the goal of early identification and intervention by increasing the number of students identified as needing supplemental instruction, therefore stressing school resources (Fletcher et al., 2002).

Some literature indicates the identification of false positives should not be a true concern (Felton, 1992). However, others have expressed a counter opinion and indicate that false positives can produce negative consequences depending on the type of decisions being made, based on the screening results. Some of the negative consequences associated with high false positive rates include wasting instructional resources (Bishop, 2003; Jenkins, 2003; O'Connor & Jenkins, 1999; Speece, 2005), dilution of instructional services for students truly in need of intensive, explicit, and systematic instruction, and unnecessarily producing parent, teacher, and/or student anxiety (Swets, Dawes, & Monahan, 2000).

In addition, Swets et al. (2000) stated that allowing an "unreasonable" number of false positives is a questionable practice, which can occur when cut-scores are lowered. In fact, the purposeful increase of classification errors is illustrated by at least two studies. Scanlon and Vellutino (1996) adjusted kindergarten screening criterion for letter naming fluency from 10 to 20 and the number of children misidentified as at risk more than tripled. Also, O'Connor and Jenkins (1999) set their first grade criteria to be sure that no child in need of supplemental instruction was missed, but the overidentification, or false positive, rate ranged from 47 to 70%.

To summarize, accurate assessment of early literacy indicator skills is not a simple process, but it is of the upmost importance because early intervention for students at risk for reading failure is so critical. There is a delicate balance between under identification (i.e., false negatives) and over identification (i.e., false positives), especially with screening tools used in the early elementary grades when floor effects are substantial. Even though the DIBELS measures have high reliability and validity rates in most areas, classification validity may be a weakness and these measures are susceptible to the problems associated with the identification of false positives.

Strategies for Reducing False Positives

Dynamic assessment. One assessment strategy that has the potential to reduce the number of false positives is dynamic assessment. Dynamic assessment (DA) is a measure of a student's potential to learn (Grigorenko & Sternberg, 1998) and is also defined as "learning potential assessment" because it can provide mediated learning that is responsive to a student's specific, identified needs (Moore-Brown et al., 2006).

DA was developed in part to limitations inherent in traditional, standardized assessment, as discussed in Chapter 1. During the administration of traditional assessments the examiner is considered an "objective" observer and does not actively intervene during testing (Caffrey et al., 2008; Haywood & Tzuriel, 2002). In addition, traditional assessment provides limited feedback and/or practice and offers no scaffolding for learning how to complete the task (Campione, 1989; Embertson, 1992; Fuchs et al., 2007). Furthermore, these assessments may not identify students who simply need more assistance with understanding the directions of the task due to a misunderstanding of directions (Campbell & Carlson, 1995; Haywood et al., 1990) and/or due to linguistic and cultural bias (Lopez, 1997).

DA may be able to help solve some of the problems discussed in the previous section related to false positives. DA is a group of approaches that are linked by the common component of building instruction and feedback into the assessment process. With these approaches, instruction and feedback are differentiated on the basis of an individual's performance on the assessment (Elliott, 2003). This type of assessment measures more than performance of skills at one point in time by allowing the examiner to administer a pretest, provide instruction related to the student's performance, and then administer a posttest (Lidz & Pena, 1996; Olswang & Bain, 1996). DA procedures attempt to change performance by offering assistance in an effort to understand a student's true learning potential (Swanson & Lussier, 2001) and can help account for variables that may underestimate an individual's ability (Haney & Evans, 1999). For example, hindrances can include unfamiliarity with the task, the language, or the materials used in traditional, standardized assessment procedures. Dynamic assessment offers the opportunity to embed instruction in the evaluation process. This assessment procedure is referred to as "dynamic" because it includes a teaching component which may change the outcome of the assessment component (Haywood et al., 1990)

One approach to DA is an interactive process that uses a test-teach-retest format to measure a particular skill. The teaching portion of the model focuses on helping the student to learn and use strategies that will help them follow the directions and understand the purpose of what is being asked (Moore-Brown et al., 2006). Several approaches can be used during the "teaching" phase of the DA process to help the student progress from skill deficiency to achievement. The examiner can offer more trials, provide information on strategies that may help the learner accomplish the task, modify the configuration of the task, and/or offer increasingly supportive prompts (Swanson & Lussier, 2001).

When comparing the performance between test and retest, significant growth indicates that the student has the ability to learn the skill and this can be a better predictor of future performance than traditional assessments (Jitendra & Kame'enui, 1993). Limited progress may indicate possible difficulties with learning and a true need for intervention (Moore-Brown et al., 2006; Tzuriel & Shamir, 2007). DA can offer several opportunities for skill attainment and student performance can be evaluated using a continuum of how readily the student learns.

Supplementing screening tools with DA procedures may help address some of the difficulties arising around the limited accuracy of screening measures. For example, it may be beneficial to supplement DIBELS subtests and other static assessments with DA, which is not susceptible to floor effects, and may help to reduce false positive rates by providing students with the extra knowledge or experience they are lacking at such a young age (Catts et al., 2009). DA may be especially useful when applied to screening measures because this method provides information about how a student will be expected to perform following classroom instruction.

One study that evaluated the effects of DA procedures on the assessment of phonological awareness skills was conducted by O'Connor and Jenkins (1999). The purpose of this study was to design a set of phonological awareness, letter, and memory tasks that could potentially identify students at risk for future reading difficulties. The researchers followed three cohorts of students (N = 445) from kindergarten through first grade over a time period of 4 years. Several measures were used to evaluate student performance throughout the investigation. For the phonological measures (i.e., syllable blending and segmenting, syllable deletion, blend phonemes, segment phonemes, first-sound isolation, rhyme production), the examiner provided instructive feedback during the assessment in order to further familiarize students with the tasks.

For students who scored less than 80% on the first test of segmentation in first grade, a dynamic segmentation task was administered following procedures developed by Slocum and colleagues (Slocum, O'Connor, & Jenkins, 1993). Each of the three trials began with a testing trial of five new words. If the student segmented fewer than four words correctly the following teaching phases were implemented: (a) model segmenting onsets and rimes and have students repeat each of the words while demonstrating with Elkonin boxes, (b) ask the student to segment five new words using the Elkonin boxes and without a teacher model, and (c) administer a trial without prompts or boxes.

Based on the results following data analysis, researchers concluded that the combination of graduated scoring and corrective feedback increased the predictive validity of the segmentation task by reducing floor effects. In addition, the procedure also reduced the false positive rate when compared to earlier prediction studies.

Another study that used DA with early literacy skills was conducted by Fuchs et al. (2007). The purpose of the study was to develop a DA measure in early reading that may be able to help school personnel identify students at risk for reading difficulties earlier and investigate the DA measure's predictive validity. One hundred and thirty-three kindergarten and first grade students met the screening criteria to be included in the study and were administered a battery of traditional tests and a DA measure focused on early literacy skills. During the next 11 weeks, students participated in reading instruction and their progress was monitored weekly. After this instructional period, the students were tested again and the data were analyzed in an effort to compare the DA measure to the other, traditional measures. The goal was to determine if the DA measure added value to the assessment protocol or whether it was simply presenting redundant results. The analysis indicated that DA was a valuable predictor of reading performance and the authors concluded that DA may help teachers (a) reduce the number of children in need of Tier 2 instruction by reducing the number of identified false positives, and (b) identify students with very low performance who will likely not respond to Tier 2 instruction and should be provided with intensive intervention at Tier 3 instead.

Although the results of these studies appear promising, DA presents some limitations making its usefulness for teachers questionable. Many of the assessment tools require a trained psychologist or diagnostician for administration. In addition, DA procedures may not be appropriate for screening purposes because they can be time consuming due to the fact that they are individualized to each student's specific needs as identified with a pretest, which prevents generalizability of treatment protocols to multiple students. For example both of the studies discussed earlier (i.e., Fuchs et al., 2007; O'Connor & Jenkins, 1999) included DA procedures that required approximately 30 min to administer with individual students.

Precorrection. Another possible solution to high false positive rates is precorrection. Compared to correction procedures that occur following an error, precorrection is proactive. It is defined as "an antecedent instructional event designed to prevent the occurrence of predictable problem behavior and to facilitate the occurrence of more appropriate replacement behavior" (Colvin et al., 1997). The specific steps that comprise the precorrection strategy were presented in Chapter 1.

The majority of research investigating the effectiveness of precorrection strategies has been in the area of behavior management (e.g., Colvin et al., 1997; Lewis et al., 2000; Oswald et al., 2005), but this strategy may also have the potential to increase academic achievement. Recent research has been conducted investigating the effects of precorrection on reading outcomes including an examination of a decoding precorrection strategy (Miao et al., 2002) and a case study that involved the preview of key words and reading passages (O'Donnell et al., 2003).

Miao et al. (2002) used a multiple-baseline design across three groups to investigate the effectiveness of precorrection used with students with mild to moderate disabilities during decoding instruction. Six students were randomly assigned to one of three instructional groups. Measures focused on reading accuracy and frequency of on-task behavior. The following procedures were used when research staff introduced the precorrection strategy to each of the experimental groups. The first precorrection strategy was *reading visually similar sounds.* The teacher identified the most difficult letters to discriminate prior to the daily lesson and modeled the correct sounds for each letter before the lesson began. The second precorrection strategy, *reading vowel sounds*, instructed students to look carefully at each vowel sound presented in the upcoming reading task. Then, the teacher modeled each vowel sound prior to students reading the list of words. Precorrection strategy three, *stopping between sounds when reading words*, began with the teacher reminding the students not to stop between sounds when they were blending sounds. The teacher then modeled how to sound out words and blend the sounds without stopping.

Results of the investigation indicated that when precorrection strategies were used in conjunction with Direct Instruction teaching methods, students' accuracy of reading sounds and words increased. During the intervention phase, the students increased their accuracy of sound identification by approximately 25% and word reading accuracy by 30 to 43%. The results also indicated that on-task behavior could be increased through the use of a precorrection strategy. The increase in percentage of on-task behaviors for the groups ranged from 14 to 28%.

Another study investigating the effects of a precorrection strategy on reading outcomes was conducted by O'Donnell et al. (2003). The precorrection strategy in this single case study was a combination of previewing passage content and discussion of key words. The student was a 10 year-old fifth grader with limited English proficiency placed in a regular education classroom setting. The authors used an ABAB reversal design to determine the effects of the intervention on the number of words read correctly and the number of correct answers to comprehension questions.

The intervention consisted of the following components completed by the experimenter (a) discussion with the student related to the target story prior to reading, (b) identification of key words in the passage, (c) modeled pronunciation of the key words and student imitation, (d) discussion of definitions for unfamiliar words and their context in the story, (e) questions to determine if the student understood the key words, and (f) the story passage was read aloud to the student. Following the model read aloud, the student was asked to read the same passage aloud and then asked five comprehension questions related to the passage. The findings indicated that the student increased the number of words read correctly and accurately answered more comprehension questions following the preview of the story and the identification of key words and their definitions.

Along with the benefits described as outcomes for the studies reviewed above, precorrection procedures may also help in the identification of false positives by helping students avoid making common errors on assessment measures. The precorrection procedures described in these studies appear practical for classroom use, but may not be appropriate for use prior to screening for some of the same reasons identified for DA. Most importantly, it would be too time consuming to precorrect errors individually for all students in a classroom.

Task Training

An approach that combines some of the features of both DA and precorrection, task training, may help to address some of the difficulties associated with using either of the interventions in conjunction with a universal screening tool. Task training is an approach that combines some of the features of both DA and precorrection. It is a brief, explicit training of the salient components of a specific task (Mackiewicz et al., 2010).

Task training includes task analysis and teaching through use of conspicuous strategies for completing a specific task. Procedures include the use of Model-Lead-Test procedures to ensure sufficient scaffolding for students who have the prerequisites needed to successfully complete the task. Task training does not include sufficient practice to be considered instructional and is therefore unlikely to boost performance on the target task if students do not have that specific skill in their repertoire. This procedure may be useful in providing a practical method for reducing the occurrence of false positives on widely used early literacy screening measures.

One study has investigated the effects of task training procedures to differentiate the need for supplemental instruction from task misunderstanding on an early literacy assessment tool (Mackiewicz et al., 2010). A randomized pretest-posttest experimental design was used to compare the effects of task training on the phoneme segmentation skills of kindergarten students. The treatment group received brief task training, while the control group received no intervention other than instruction within the general education classroom. Significant differences were found between the two groups on a posttest measure of phoneme segmentation fluency. However, only one early literacy skill, phoneme segmentation, was targeted during that investigation. As discussed earlier, most schools use a combination of four DIBELS subtests at midyear kindergarten to determine the need for supplemental instruction. Therefore, future research with task training should investigate its effects on the combination of early literacy measures used to identify kindergarten students in need of supplemental instruction.

Summary of Research

Extensive research supports early intervention and most reading difficulties, and even reading disabilities can be prevented when intervention begins in kindergarten or first grade. One model of early identification and intervention is RTI and preliminary evidence indicates that RTI may improve reading outcomes for students, reduce the number of special education referrals and placements, and reduce disproportionality of males placed in special education programs. All of these benefits lead to a reduction in costs for school districts because the funds allocated for conducting evaluations can be redirected to assist with prevention efforts and/or direct intervention.

The success of early intervention models, including RTI models, hinge of the accurate assessment and identification of students at risk for reading failure. Accurate assessment is necessary in order to deliver effective early intervention. High false positive rates, resulting from low cut-score and floor effects, may interfere with the effectiveness of early intervention programs by stressing school resources and diluting what should be intensive intervention.

Dynamic assessment and precorrection strategies, when combined with screening administration, show promise in reducing false positive rates. However, these methods may be too time consuming and individualized to effectively generalize to large groups of children typically assessed with universal screening measures. Further investigation is needed to determine if an approach that combines features of DA and precorrection, task training, reduces false positive rates. Task training has the potential to lead to more informed, and thereby more accurate, instructional decision making for students at risk, but more research is needed that includes task training procedures for several predictors (i.e., phonemic awareness, rapid letter naming, alphabetic understanding).

CHAPTER 3: METHOD

This study investigated the effects of task training, targeting three early literacy measures, in order to differentiate the need for supplemental instruction from task misunderstanding for kindergarten students identified as at risk for future reading difficulties. This chapter presents the method used to investigate the research questions including information describing participants, instrumentation, data collection procedures, task training procedures, research design, and a description of the data analyses that was conducted.

Participants

All kindergarten students enrolled in the participating school served as the group from which the final 42 participants were selected. Consent letters were sent to parents of all kindergarten students at the school. The consent form explained the study procedures and asked parents for permission for their child to participate in the study.

Next, pretests were administered to all students whose parents consented to allow them to participate (n=60). The pretest consisted of the DIBELS subtests appropriate for administration during the kindergarten midyear benchmark. Results of the pretest were used to determine each student's instructional status recommendation, "*Benchmark At grade level,"* *"Strategic – Additional intervention," or "Intensive – Needs substantial intervention."* Students whose instructional recommendation was either ***strategic*** (n=30) ***or intensive*** (n=12) continued as participants in the study. Prior to beginning

Intervention, the students were randomly assigned into the treatment and control groups. Demographic information regarding the student participants are shown in Table 1.

Table 1. *Demographic information for participants*

	N	Percent
Group		
Treatment	20	47.6
Control	22	52.4
Gender		
Male	23	55
Female	19	45
Ethnicity		
African American	38	90.4
Asian	2	4.8
Hispanic	2	4.8

In order to collect social validity data regarding treatment acceptability, 6 kindergarten teachers who taught at the participating school, with a history of administering and scoring the DIBELS, were asked to complete a questionnaire.

Setting

An elementary school in an urban school district in the southeast United States was used as the setting for this study. The school was selected based on the sufficient number of kindergarten students for conducting a group comparison study, the socioeconomic status level of the school population, and the principal's willingness to have the school participate in the study. According to information from the 2008-2009 school year, the school serves approximately 577 students in kindergarten through fifth grade. Demographic information related to the school's total population indicates that 90.8% are African American, 4.2% are Hispanic, 2.6% are Multi-Racial, 1.4% are Caucasian, and 1% are Asian American. The school has 297 (51.5%) male students and 280 (48.5%) female students. With regards to socioeconomic status, approximately 93% of the students qualify for free or reduced-price lunch.

Pretest and posttests were administered in quiet place in the hallway outside the kindergarten classrooms with individual students. Task training was completed in groups of two or three participants in a tutoring room designated for use by tutors working with small groups. Task training occurred before and after the tutoring groups were conducted. Interventionists and participants were seated at rectangular-shaped tables, on opposite sides of the table.

Researcher

The researcher is a doctoral candidate in Special Education in the Department of Special Education and Child Development at the University of North Carolina at Charlotte. She has Bachelor of Arts degree and a Master of Education degree in Special Education. She has 2 years experience teaching students with emotional and behavioral disabilities within a separate setting. She then earned a Specialist in School Psychology degree and has received training in the administration and scoring of the DIBELS subtests. The researcher has been a practicing school psychologist in both prekindergarten and elementary schools for 7 years. Her role in the study was to (a) administer pretest and posttest measures, (b) provide training to an additional assessor and interventionist, (c) implement task training, and (d) analyze data.

Interventionists and Assessors

The researcher served as an interventionist and assessor. An additional interventionist and assessor was selected from interested UNC Charlotte graduate students majoring in special education, recommended by a special education faculty member, and was available at the scheduled times. The researcher and the graduate student conducted all assessments and provided task training to the participants in the treatment group.

Prior to initiation of the study, the researcher followed several steps in training the graduate student to conduct the intervention. First, the graduate student viewed video recorded task training sessions. Next, the graduate student practiced the task training procedures with the researcher. Finally, training included practice with kindergarten students not included in the study until 90% or greater accuracy was reached on the treatment fidelity checklist for each measure.

The same graduate student assisted the researcher with data collection after being trained in the administration of the four DIBELS subtests. The researcher completed an administration fidelity checklist for each of the four measures and scored student responses along with the graduate student. Training continued until the graduate student completed all assessment steps with 100% accuracy and when at least 90% agreement was reached between the researcher and graduate student in an item-by-item analysis on each of the measures. The researcher and the trained graduate student administered all measures.

Independent Observer

An independent observer collected treatment fidelity data and interscorer reliability data. The observer remained naïve to the purpose of the study and any expectations regarding the outcomes. Prior to beginning the study, the independent observer was selected from interested UNC Charlotte undergraduate students majoring in special education, recommended by a special education faculty member, and was available at the scheduled times.

The observer was trained to record data on each of the treatment fidelity checklists (see Appendix B) used when observing the task training interventions. During the observer training, interventionists (researcher and graduate student) and the independent observer simultaneously completed treatment fidelity checklists for videotaped training sessions. A total of six task training sessions were observed (two for each subtest). Interobserver agreement was calculated and brought to a level of 100% agreement prior to initiation of the study.

The observer also was trained on the administration and scoring procedures of the four DIBELS subtests administered during the study using videotaped subtest administrations and in vivo role play by the assessors acting as administrator and student. Training sessions included multiple opportunities to practice scoring, refine scoring procedures, and reconcile discrepancies between the assessors and the second observer. Training continued until interscorer agreement reached at least 90% with both assessors on each subtest.

Instrumentation

The four DIBELS subtests used to establish risk level for kindergarteners at the midyear benchmark period (i.e., Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency) were used for pretesting and posttesting.

Initial Sound Fluency (ISF). The ISF subtest measures a student's ability to recognize and produce the initial sound in an orally presented word, and therefore, is a measure of phonological awareness (Good, Laimon, Kaminski, & Smith, 2002). The child is presented with four pictures; the examiner names each picture, and then prompts the children to identify, by pointing or saying, the picture that begins with a sound produced orally by the examiner. For example, the examiner will say, "This is road, barn, hand, and egg. Which picture begins with /b/?" and the student should point to the picture of the barn or say "barn." The examiner will also ask the child to say the beginning sound for an orally presented word that matches one of the pictures. For example, the examiner will ask, "What sound does egg start with?" and the child should answer by saying "/e/." The amount of time taken to complete all questions will be calculated and converted into the number of onsets correct per minute.

The ***established*** goal for the ISF subtest in middle of kindergarten is 25 initial sounds correct per min. Students scoring between 10 and 24 correct sounds are considered to have ***emerging*** skills, and students scoring fewer than 10 correct sounds are considered to have a ***deficit*** in this area.

The ISF measure is a revision of a previous subtest, Onset Recognition Fluency (OnRF), with minimal changes. Alternate-form reliability of the OnRF measure is .72 in

January of kindergarten (Good et al., 2004). When the assessment is repeated four times, the average is estimated to have a reliability of .91 (Nunnally, 1978). Good et al. report the concurrent, criterion-related validity of OnRF with PSF is .48 in January of kindergarten and .36 with the Woodcock-Johnson Psycho-Educational Battery (WJ-III ACH; Woodcock et al., 2000)-Readiness Cluster score. The predictive validity of OnRF with respect to spring of first grade reading on CBM ORF is .45 and .36 with the Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster score (Good et al.). In addition, Elliot et al. (2001) investigated the reliability and validity of modified versions of the DIBELS subtests. The concurrent criterion-related validity of the modified ISF measure ranged from .42 to .64. The reliability ranged from .64 to .89. However, these findings are limited because the modified version of ISF was accuracy based, as opposed to the fluency-based design of the original DIBELS subtests, which will be used in the current study.

Letter Naming Fluency (LNF). This subtest provides a measure of risk, but does not directly correspond to any of the big ideas of early reading; however, rate of naming letters has been found to be a strong predictor of later reading performance (Burke et al., 2009; Rouse & Fantuzzo, 2006; Stage et al., 2001). Students in the lowest 20% of a school district are considered ***at risk*** for future reading difficulties and students scoring between the 20th percentile and 40th percentile are considered to be at ***some risk***. To administer this subtest, the examiner presents the student with a page of letters, both upper and lowercase, arranged in random order. Students are asked to name as many letters as they can and will be told that they will be told the letter name if they do not know it. The student will name as many letters as he or she can within 1 min and the total number of letters correctly named within this time will be the student's score (Kaminski & Good, 2002).

The ***low risk*** goal for the LNF subtest in middle of kindergarten is 27 letter names correct within 1 min. Students scoring between 15 and 26 correct letter names are considered to be at ***some risk***, and students scoring less than 15 correct letter names are considered to be ***at risk*** in this area.

The one-week alternate form reliability of LNF is .93 (Kaminski & Good, 1996), while the one-month alternate-form reliability of LNF is .88 in kindergarten (Good et al., 2004). The median criterion-related validity of LNF with the Woodcock-Johnson Psycho-Educational Battery (Woodcock et al., 2000) – Revised Readiness Cluster standard score is .70 in kindergarten. The predictive validity of kindergarten LNF with first grade Woodcock-Johnson Psycho-Educational Battery – Revised Reading Cluster standard score is .65 and .71 with first-grade Curriculum-Based Measurement oral reading fluency (Good et al.). In addition, Burke et al. (2009) found the validity of LNF when predicting Oral Reading Fluency scores in second grade to be .63.

Phoneme Segmentation Fluency (PSF). The PSF subtest has been found to be a strong predictor of later reading achievement (Good, Kaminski, & Smith, 2002) and is a standardized, individually administered test of phonemic awareness. This subtest assesses the student's ability to fluently segment words with two to five sounds into their individual phonemes. The PSF task is administered through the oral presentation of words and the student is required to verbally produce the individual phonemes for each word. For example, if the examiner says "bat," the student must say "/b/ /a/ /t/" in order to receive all three possible points for that word. The number of correct phonemes produced by the student after 1 min will determine the student's final score.

The ***low risk*** goal for the PSF subtest in middle of kindergarten is 18 phonemes correct within 1 min. Students scoring between 7 and 17 correct phonemes are considered to be at ***some risk***, and students identifying fewer than 7 correct phonemes are considered to be ***at risk*** in this area.

Various reports have demonstrated technical adequacy related to reliability, validity, and sensitivity to small changes in skill acquisition for the PSF subtest. The two-week, alternate form reliability for the PSF measure is .88 (Kaminski & Good, 1996), and the one-month, alternate-form reliability is .79 in May of kindergarten (Good et al., 2004). The concurrent, criterion-related validity is .54 with the Woodcock-Johnson Psycho-Educational Battery Readiness Cluster (Woodcock et al., 2000) score in spring of kindergarten (Good et al.). The predictive validity of spring-of-kindergarten PSF with (a) winter of first grade DIBELS NWF is .62, (b) spring of first grade Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster (Woodcock et al.) score is .68, and (c) spring of first grade CBM ORF is .62 (Good et al.).

Nonsense Word Fluency (NWF). The DIBELS NWF measure is a standardized, individually administered test of the alphabetic principle that assesses a student's ability to blend letters, representing their most common sounds, into words (Kaminski & Good, 1996). On the NWF measure the student is presented with an 8.5" x 11" sheet of paper with randomly ordered vowel – consonant (VC) and consonant – vowel – consonant (CVC) nonsense words (e.g., vum, et, bec) and the examiner asks the student to say the individual letter sounds or read the whole nonsense word. The student will be allowed 1 min to produce as many letter sounds as he or she can. The final score will be the number of correct letter sounds produced within the minute.

The ***low risk*** goal for the NWF subtest in middle of kindergarten is 13 correct letter sounds within 1 min. Students scoring between 5 and 12 correct letter sounds are considered to be at ***some risk***, and students scoring fewer than 5 correct letter sounds are considered to be ***at risk*** in this area.

The one-month, alternate-form reliability for NWF in January of first grade is .83 and the concurrent criterion-validity with the Woodcock-Johnson Psycho-Educational Battery-Revised Readiness cluster (Woodcock et al., 2000) score is .36 in January and .59 in February of first grade. The predictive validity of the measure in January of first grade with (a) CBM ORF in may of first grade is .82, (b) CBM ORF in May of second grade is .60, (c) Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster (Woodcock et al.) score is .66 (Good et al., 2004).

Interscorer Reliability

Interscorer reliability was calculated from data collected by the second observer when compared to data collected by the assessors. Agreement data was collected by the independent observer using direct observation of 31.5% of the pretests and 28.6% of the posttests. Item-by-item agreement (Tawney & Gast, 1984) was recorded by the researcher following each observation. An agreement was recorded if both the assessor and observer identically scored the item as correct or incorrect. A disagreement was recorded if the task was not scored identically. The researcher calculated percent agreement for each task by dividing the number of agreements by the number of agreements plus disagreements multiplied by 100. Interscorer reliability data is reported in Chapter 4.

$$\frac{\text{agreements}}{\text{agreements + disagreements}} \quad \text{X 100 = percent agreement}$$

Dependent Variables

The primary dependent measure for this study was change in instructional status recommendation for participants. The recommendation is determined following entry of student scores on the four DIBELS subtests into the DIBELS Data System. A comparison was made between the treatment and control groups on the percentage of students at each level (***benchmark***, ***strategic***, ***intensive***) for each test (pretest, posttest).

Three other dependent variables were measured in the study. First, data related to the number of initial sounds isolated per min on an initial fluency task were collected. Another dependent variable measure was the number of correct phonemes the student identified within 1 min when orally presented two to five phoneme words. The third additional measure will be the number of sounds in nonsense words correctly read within 1 min. A comparison of the two groups' posttest performance on each of the subtests was completed.

Procedures and Data Collection

This section describes the general procedures and data collection procedures that were followed including pretest administration, participant selection, task training procedures, treatment fidelity, posttest administration, and social validity. In addition, the timeline for data collection will be described.

Pretest. The pretest consisted of the four DIBELS subtests administered at the midyear benchmark in kindergarten (i.e., ISF, LNF, PSF, NWF). Subtests were administered to individual students in a quiet place in the hallway outside the kindergarten classrooms with individual students. Students were encouraged to do their best and also told that it was okay if they did not know all of the answers.

On the ISF subtest, the student was asked to identify a picture from an array of four pictures that begins with a sound orally produced by the examiner. The examiner named each picture and prompted the student to identify, by pointing or saying, the picture that begins with a sound produced orally by the examiner. The procedure was repeated for three of the four pictures. Then, the student was asked to give the beginning sound for an orally presented word that matches one of the pictures. After each question was asked, the examiner started a stopwatch and then stopped the stopwatch as soon as the student finished their response. This sequence was repeated across four sets of pictures for each student.

Student responses were scored as either correct (1 point) or incorrect (0 points). If the student did respond after 5 s, the question was scored as zero and the next question was presented. When the student finished the last question, the total time on the stopwatch was recorded in seconds and the number of correct responses was counted. The ISF score (number of onsets correct per min) was calculated using the formula:

$$\text{ISF} = \frac{\text{60 x number correct}}{\text{seconds accumulated during the test}}$$

According to the assessment developers, this measure takes approximately 3 min to administer and has over 20 alternate forms.

To administer the LNF subtest, the examiner presented the student with a page of letters, both upper and lowercase, arranged in random order. Students were asked to name as many letters as they could and were told that they would be told the letter name if they did not know it. After the directions had been given to the student, the stopwatch was started. The student named as many letters as they knew and at the end of 1 min, the examiner told the student to stop. If the student did not get any correct letter names within the first 10 letters (one row), the task was discontinued and a score of zero was recorded. If the student hesitated for 3 s on a letter, the letter was scored as incorrect and the letter name was provided. The student's score was the number of letters named correctly in 1 min.

The PSF task was administered through the oral presentation of words and the student was required to verbally produce the individual phonemes for each word. After the directions were given to the student and one practice item had been completed, the examiner orally produced the first word and the stopwatch was started. If the student did not say a sound segment after 3 s, the second word was given and the first word was scored as zero segments produced. As soon as the student finished saying the sounds, the next word was presented promptly and clearly. If the student did not give produce phonemes correctly in the first 5 words, the task was discontinued and a score of zero was recorded. At the end of 1 min, presentation of words stopped. The number of correct phonemes produced by the student after 1 min determined the student's final score for this subtest. The PSF measure takes approximately 2 min to administer and has over 20 alternate forms.

On the NWF measure the student was asked to say the individual sound of each letter or read the whole nonsense word. Following the directions, the student was told to begin, and the stopwatch was started. If the student did not get any sounds correct in words 1 through 5, the task was discontinued and a score of zero was recorded. The student was allowed 1 min to produce as many letter sounds as he or she could produce. The final score was the number of correct letter sounds produced within 1 min. The NWF measure takes approximately 2 min to administer and has over 20 alternate forms.

Participant selection. Following the pretest administration, all scores were entered into the DIBELS Data System (www.dibelsuoregon.edu) in order to obtain an instructional status recommendation. The DIBELS Data System is a web-based database that schools can use to enter student performance results and create reports based on the scores. The system was developed in 2001 and is maintained by personnel at the Center on Teaching and Learning (CTL) affiliated with the University of Oregon. During the 2007-2008 school year, the DIBELS Data System was used in over 15,000 schools. It is a fee-based service with the cost being $1 per student per academic year.

After all of the scores were entered into the database, an instructional status recommendation was calculated for each student. Those students whose instructional recommendations were either ***strategic*** or ***intensive*** continued as participants in the study. These students were randomly assigned to either the task training group or the control group. Computer generated random assignment was used.

Task training. Task trainings for three of the four subtests was provided for students in the treatment group. The task of naming letters on the LNF subtest is simpler and the directions are clearer than for the other subtests, so that it is anticipated that errors would be less likely to occur due to task misunderstanding. The most likely error, saying the letter sound rather than the letter name is addressed during the administration procedures. The examiner is allowed to give the following prompt one time during the administration: “Remember to tell me the letter name, not the sound it makes.” Task training does not appear to be as necessary for this measure so it was not included in the current study.

Students in the treatment group only participated in task training for those subtests on which their performance was classified as being at ***some risk*** or ***at risk*** on the PSF and NWF measures and as being within the ***emerging*** and ***deficit*** ranges on the ISF measure. For example, if a student was identified as ***at risk*** on PSF and NWF, but not on ISF, that student only participated in task trainings for PSF and NWF. In an effort to lessen order effects, students in the treatment group were randomly assigned to three groups using computer-generated assignment procedures. Each of the three groups participated in the task trainings in a different order. Table 2 displays the order in which the groups participated in the task training sessions. Students in each group were then randomly assigned to small groups for intervention.

Table 2. *Order of task training administration*

	Training 1	Training 2	Training 3
Group A	ISF	PSF	NWF
Group B	PSF	NWF	ISF
Group C	NWF	ISF	PSF

Students assigned to the treatment group were trained in groups of two or three. ISF task training sessions ranged from 6 min to 8 min with a mean of 7 min. Task training sessions for PSF ranged from 7 min to 14 min with a mean of 10 min. NWF task training sessions ranged from 9 min to 15 min with a mean of 12 min. Variablility in training length occurred as a result of the number of error corrections needed by each group.

A model, lead, test, feedback strategy (i.e., my turn – together – your turn) was used to teach students how to correctly respond to subtest expectations. The task training scripts were designed to familiarize students with the task directions, expose them to the specific language used by examiners, provide group and individual practice, and correct common errors. For all of the task trainings, interventionists followed a script and recorded the length of the task training session. The series of three task training scripts are provided in Appendix A.

The ISF task training used a model-lead-test format to teach students how to complete each of three tasks. The first task is Picture Naming. The interventionist gave each student an 8 ½” x 11” piece of paper with four color pictures on it and modeled naming the pictures while pointing. The students and the interventionist named each picture while pointing. Then, all of the students named each picture in unison while pointing, then each student was given a turn to name all of the pictures independently.

The next task is Identifying Initial Sounds. First, the interventionist modeled saying the name of the object in the picture and producing the initial sound. For example, while pointing to a picture of *pig,* the interventionist said, “I will say the name of the picture and then tell if it begins with /p/. Listen. *Pig*, /p/.” The interventionist then pointed to a picture of a *flower* and said, “*Flower* does not begin with /p/.” Then the interventionist pointed to the remaining pictures while saying the name of the object in the picture and emphasizing the beginning sound. Next, students were led by the interventionist to identify the objects in the pictures. The students were asked a series of questions about the pictures including “Does *sock* begin with /s/?” The students produced the target (/s/) sound and point to the picture that begins with /s/. These procedures will be repeated with two pictures. In the last step, the interventionist will point to pictures and ask questions about what the objects in the pictures are while the students answer. For example, the interventionist will say “What is this?” while pointing to a *bus.* After the students answer, the interventionist will ask “Does *bus* begin with /b/?” Then, students will say or point to the correct pictures when asked the question “Which picture begins with /__/?” After the students answer the questions in unison, each student will be given an opportunity to do one item individually.

The third task in this task training is Producing Initial Sounds. First, the interventionist modeled the skill saying “My turn. I will say the sound that *saw* begins with - /s/. Listen, /s/, *saw*." Then the interventionist will ask, "What sound does *saw* begin with?" and the students will answer /s/ in unison. Next, the interventionist and the students will answer a question about each word represented by a picture. For example, the interventionist will say, "*Cow* begins with /k/. What sound does *cow* begin with?" and both the interventionist and the students will answer in unison. In the last step, each student will be given an opportunity to identify the initial sound in a word.

Error correction procedures followed a model-lead-test format and are included in the script. When an error occurs, the interventionist immediately began the error correction procedures. The interventionist followed the specific steps for the particular task where the error occurred (see Appendix A). Error correction procedures were administered to the whole group even on individual turn errors.

The PSF task training also used a model-lead-test format to teach students how to complete the task. The first step of the training was the model. Students were shown a picture of a *sun* and asked to name the picture. Then, the interventionist modeled saying the word the "fast way" (i.e., at normal word reading rate) and told students that each sound in the word can also be said. Individual sounds (i.e., /s/-/u/-/n/) were said aloud while the interventionist raised one finger for each sound. While fingers were still raised, the interventionist asked the group, "How many sounds are in *sun*?" Next, the interventionist modeled how to tap the table one time for each sound in the word. These procedures were followed for one additional word, *ice*.

The next step in the PSF task training procedures were the lead. Students were shown another picture and asked to name it. The interventionist segmented the word raising one finger for each sound that was said. Students were asked "How many sounds are in *book?"* and the students answered in unison. Then, the examiner asked "How many times will we tap the table?" and again the students answered in unison. The interventionist and students tapped the table one time for each sound while saying the sounds in the word *book.* These procedures were repeated with one additional word, *fan.*

Then, the procedures for the lead step were shortened. The interventionist said "Everybody, get ready to tap the table one time for each sound. Tell me the sounds in *eat.*" The interventionist and students said the sounds in unison while tapping the table. These abbreviated lead procedures were repeated with one additional word, *mom.*

The next step in the PSF task training was the test. The interventionist said "Tell me the sounds in *zoo"* and all of the students said the sounds while tapping the table. This step was repeated with the words *sit*, *if*, and *fun*. Individual tests were then given to the students. The interventionist said a child's name and "Your turn to tell me the sounds in *no.*" Individual turns were given until each student has had an opportunity to practice segmenting at least two words.

Error correction procedures also followed a model-lead-test format and are included in the task training script. When an error occurs, the interventionist immediately began the error correction procedures. The interventionist modeled saying each sound in the word while tapping. Then, the students and interventionist said each sound while tapping. Finally, the students tapped alone and said the sounds in the missed word. Error correction procedures were administered to the whole group even on individual turn errors.

The NWF task training used model-lead-test and model-test formats to familiarize students with task expectations. There were five phases of the training. Each student was given a “student sheet” for use during the training. The 8 ½” x 11” piece of paper had seven rows of items on it, ranging from individual letters (consonants and vowels) to two- and three-phoneme nonsense words.

In the first phase, Identifying Consonant Letter Sounds, the students were directed to the first row of letters. The interventionist modeled saying the sound for each corresponding letter, while the students pointed to the letter. Then, the students said each sound in the first row in unison while pointing to the corresponding letter.

The second phase, Identifying Short Vowel Letter Sounds, provided an opportunity for students to practice saying short vowel sounds, which is the expectation during this particular subtest. The interventionist modeled saying each short vowel sound while the students pointed to each letter. Then, the students were given an opportunity to produce the short vowel sounds in unison while pointing.

The third phase, My Turn – Your Turn: Reading Each Word, moved from identifying letter sounds to blending sounds together to make “make-believe words.” First, the interventionist explained the term “make-believe words” and modeled saying the individual sounds of the nonsense words while pointing to each letter. The interventionist modeled the sounds in the first word and the students said the sounds in the first make-believe word in unison. This procedure was repeated for the remaining four words in the row.

The fourth phase of the training will be My Turn – Your Turn: Reading Row of Words. In this task, the interventionist modeled saying the sounds in each make-believe word in the fourth row. The students followed along with their fingers. Then, the students said the sounds in all of the words in the same row. The last phase of this task training will focus on Individual Turns. Individual students said the sounds in the make-believe words for one row (five words). Students were reminded that they could skip letters they did not know.

Error correction procedures followed a model-lead-test format as well and were included in the task training script. For rows 1 through 4, error correction procedures were administered to the whole group. For rows 5 through 7, students said the sounds or read the words individually and the scoring rules for the NWF subtest were followed. No error correction was provided for missed sounds on these rows. Only procedural errors (e.g., not skipping letters they do not know, saying the long vowel sound, substituting real words for make-believe words) were corrected.

Treatment fidelity. A series of three treatment fidelity checklists corresponding to the three task training scripts were used to ensure the consistency of task training implementation across interventionists and to ensure that the intervention was being implemented as designed. See Appendix B for each of the treatment fidelity checklists. The independent observer directly observed 27.3% of the task training sessions, evenly distributed across interventionists and across the three interventions. The observer recorded whether the interventionist presented each step of the task training.

The interventionists and observer recorded the length of the task training sessions and the observer recorded whether the interventionist adhered to the script and procedures for each of the trainings. As described earlier, the task training scripts are individualized to each specific subtest, so the treatment fidelity checklists are organized differently from each other.

Although the task trainings are scripted, minor deviations from the script were acceptable. An interventionist can slightly change the wording of a task, as long as the modification does not affect the task. For example, the scripted text reads “I will tap the table ___ times while saying the sounds in the word, one tap for each sound.” If the interventionist said “I can tap the table ___ times as I say the sounds in the word, a tap for each sound.” The change in wording does not affect the task; therefore, the change is acceptable. Conversely, major deviations (e.g., changing wording that alters the task, providing extraneous information for more than a sentence, or leaving out components of the training) were not acceptable. Next, the treatment fidelity checklists will be described.

The ISF task training treatment fidelity checklist contained 46 items. The PSF checklist contained 33 items, while the NWF checklist contained 62 items. On each checklist, every step of the specific task training is listed and the observer indicated if that step was completed by circling the word “YES” on the checklist. If the step was omitted by the interventionist, then the observer circled the word “NO.” If a step was not needed during the training, then “NA” was circled on the form. For example, if no errors were made during the lead step of a particular task training, then “NA” was circled for each of the error correction steps.

Of the six task training sessions that were observed for treatment fidelity purposes, three (50%) were video recorded for the purpose of collecting reliability data. A third observer completed treatment fidelity checklists while viewing the video recorded sessions. Item-by-item agreement (Tawney & Gast, 1984) was recorded by the researcher following the observations. An agreement was recorded if the second and third observers identically scored an item by circling “YES,” “NO,” or “NA.” All treatment fidelity data is discussed in Chapter 4.

Posttest. Students in both the treatment and control groups were administered a posttest, beginning one school day following the conclusion of all three of the task training sessions. The posttest consisted of alternate forms of each of the four DIBELS subtests including ISF, LNF, PSF, and NWF. All participants were administered the four subtests, even if they did not participate in task training for a particular subtest. Posttesting followed the same procedures described earlier for pretesting.

Social validity. At the conclusion of the study, 6 kindergarten teachers at the participating school, experienced in the administration of the DIBELS measures, participated in a social validity session designed to determine their acceptability of the intervention and to glean their opinions about the feasibility of the task training procedures. Social validity data was collected prior to the results of the study being shared with school personnel. The same person who served as the third observer for reliability of treatment fidelity during the study facilitated the social validity session. First, teachers were thanked for sharing their students for the purposes of the study and were told that the task training procedures would be shared with them that day in order to learn more about how they viewed the trainings in terms of practicality, usability, and potential for teachers to use to help prepare students for certain assessments. The teachers were each given a copy of the social validity questionnaire and watched a video-recorded demonstration of ISF task training and, at the conclusion of the demonstration, were asked to answer question one. The same procedure was followed for the PSF task training with question two and the NWF task training with question three. Then, the teachers were asked to consider what they thought about the need for a task training to address the LNF subtest and answer question four. Upon completion of question four, the teachers were asked to complete the remaining questions on the form. Questionnaires were analyzed based on themes that emerge from teacher responses. The questionnaire is included in Appendix C and an analysis of teacher responses will be shared in Chapter 4.

The timeline for specific pretest and posttest data collection is presented in Table 3. Pretesting occurred over three school days, as did the task training sessions. Posttesting occurred over a two-day period.

Table 3. *Timeline for data collection*

Procedures	Study Days
Pretest Administration	Days 1, 2, 3
Task Training Group A_ISF Group B_PSF Group C_NWF	Day 4
Task Training Group A_PSF Group B_NWF Group C_ISF	Day 5
Task Training Group A_NWF Group B_ISF Group C_PSF	Day 6
Posttest Administration	Days 7, 8

Research Design

A randomized controlled trial was used to investigate the effects of task training on the performance of kindergarten students identified as at risk for future reading difficulties on four early literacy indicator measures. Based on the results of a pretest, instructional status recommendations were made for students and these recommendations included ***benchmark***, ***strategic***, or ***intensive*** according to the benchmark goals identified by the DIBELS developers. Students with a recommendation of ***strategic*** or ***intensive*** were randomly assigned to either the task training group or control group.

Data Analysis

This section describes the procedures that were used in analyzing the data in order to address the research questions. An experimental, randomized design was used in this study. The first research question was examined using descriptive statistics to analyze the performance of the two groups across the two tests (i.e., pretest, posttest). Percentage of students at each level (i.e., ***benchmark***, ***strategic***, ***intensive***) were calculated and compared across groups and tests. In addition, to assess whether there was a difference in status level changes between the two groups, status change and group were analyzed using a chi-square test.

Research questions two through four were investigated first using a multivariate analysis of variance (MANOVA). The procedure simultaneously compared the independent variable (group: treatment and control) across the three dependent variables including the change score between pretest and posttest for the (a) number of initial sounds correctly isolated, (b) number of phonemes correctly identified, and (c) number of letter sounds correctly identified. Finally, one-way Analysis of Variance (ANOVA) was performed for each dependent variable in order to examine differences between the groups.

CHAPTER 4: RESULTS

The purpose of this study was to examine the effects of task training, targeting three early literacy measures, in order to differentiate the need for supplemental instruction from task misunderstanding for kindergarten students identified as at risk for future reading difficulties. This chapter will present the results for each of the research questions.

Treatment Fidelity

Procedural reliability data were collected for 27.3% of the task training sessions by an independent observer using treatment fidelity checklists developed for each task training. The task training scripts are individualized to each specific subtest, so the treatment fidelity checklists are organized differently from each other. On each checklist, every step of the specific task training is listed and the second observer indicated that a step was completed by circling the word “YES” and circled the word “NO” if the step was omitted by the interventionist. If a particular step was not needed during the training, then “NA” was circled. For example, if no errors were made during the lead step of a particular task training, then “NA” would be circled for each of the error correction steps. Observations were equally distributed between the two interventionists and among the three task training procedures. Overall treatment fidelity was rated 96.5% (range 90.2% to 100%).

For the Initial Sound Fluency task training, a 46-item checklist (see Appendix B) was used to measure the integrity of delivering this particular task training to a group of students. The independent observer recorded the length of the instructional sessions as well as whether the interventionist (a) adhered to the script and procedures for the Picture Naming phase, (b) adhered to the script and procedures for the Identifying Initial Sounds phase, (c) adhered to the script and procedures for the Producing Initial Sounds phase, and (d) used the error corrections procedures specified in the script. Results indicated that the Initial Sound Fluency task training was implemented with a mean accuracy of 98.9% (range 97.8% to 100%).

For the Phoneme Segmentation Fluency task training, a 33-item checklist (see Appendix B) was used to measure the integrity of delivering this particular task training. The independent observer recorded the length of the instructional sessions as well as whether the interventionist adhered to the script and procedures including use of (a) model, (b) lead, (c) test, and (d) error corrections if applicable. Results indicated that the Phoneme Segmentation Fluency task training was implemented with a mean accuracy of 98.0% (range 96.0% to 100%).

For the Nonsense Word Fluency task training, a 62-item checklist (see Appendix B) was used to measure the integrity of delivering this particular task training to a group of students. The independent observer recorded the length of the instructional sessions as well as whether the interventionist adhered to the script and procedures for the (a) Identifying Consonant Letter Sounds phase, (b) Identifying Short Vowel Letter Sounds phase, (c) Reading Each Word phase, (d) Reading Row of Words phase, and (e) Individual Turns phase. In addition, it was also noted whether the interventionists provided feedback and/or error corrections as specified in the script. Results indicated that the Nonsense Word Fluency task training was implemented with a mean accuracy of 92.5% (range 90.2% to 94.8%).

Of the six task training sessions that were observed for treatment fidelity purposes, three (50%) were video recorded for the purpose of collecting reliability data. A third observer completed treatment fidelity checklists while viewing the video recorded sessions. Item-by-item agreement (Tawney & Gast, 1984) was recorded by the researcher following the observations. An agreement was recorded if the second and third observers identically scored an item by circling “YES,” “NO,” or “NA.” The mean reliability was 94.47% (range 90.9% to 100%).

Assessment Integrity and Interscorer Reliability

All assessment integrity and interscorer reliability data was collected by an independent observer. Direct observation was used by the observer to complete integrity checklists and score 31.5% of the pretests (four subtests) and 28.6% of the posttests (four subtests) administered to the participants. The percentage of tests observed was equally distributed between the two test administrators.

The DIBELS Assessment Integrity Checklists (Good & Kaminski, 2007) were used to determine if each subtest was administered in a standardized manner. As the observer watched the administration, a “✓” was recorded under the “Fine” column if the test administrator completed a step correctly and a “✓” was placed in the “Needs Practice” column if the step was not completed or completed incorrectly. Following the observations, the number of times that a “✓” was placed in the “Fine” column was calculated and divided by the total number of steps on the Assessment Integrity Checklist. The product was multiplied by 100, resulting in the percentage of steps correctly implemented during the subtest administration. Overall for the pretest, a mean of 97.2% (range 98.2% to 99.0%) of test administration steps were correctly completed by the assessors. For the posttest, a mean of 97.7% (range 97.0% to 98.3%) of the steps were completed correctly.

For interscorer reliability, an agreement was counted if both the test administrator and the observer marked the same test item as correct or the same test item as incorrect. A disagreement was counted if the second observer's markings differed from those of the test administrator. Interscorer reliability was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100 (Tawney & Gast, 1984). Overall, the mean pretest reliability was 94.9% (range 89.9% to 97.9%) and the mean posttest reliability was 94.4% (range 88.6% to 97.9%). Mean reliability data for the pretest and posttest across each subtest is included in Table 4.

Table 4. *Mean percentage of interscorer reliability across subtests*

	ISF	LNF	PSF	NWF
Pretest	96.2	97.9	89.9	95.5
Posttest	97.9	97.3	88.6	93.8

Effects of Task Training on Recommendation Status

Will there be a significant difference between students who receive task training and students who do not receive training on the DIBELS instructional status recommendation for supplemental instruction?

Overall performance on the four DIBELS subtests leads to an instructional status recommendation for each student (i.e., ***intensive***, ***strategic***, ***benchmark***). Descriptive statistics were used to examine the percentage of students categorized in each group based on their pretest and posttest performance. Table 5 presents the percentages at each status recommendation level for pretest and posttest for each group.

Table 5. *Percentages at each risk level across measures*

	Pretest	Posttest
Treatment Group		
Benchmark	0	65%
Strategic	75%	20%
Intensive	25%	15%
Control Group		
Benchmark	0	18%
Strategic	73%	59%
Intensive	27%	23%

Score difference comparisons were made between the pretests and posttests and differences were examined for each group. Following the pretest, due to the study's design, all participants' status recommendation was either ***strategic*** or ***intensive***. For the treatment group, 75% of students' performance was classified as needing ***strategic*** support and 25% was classified as needing ***intensive*** support following pretest administration. Following task training for the treatment group, only 20% of these students were still classified as needing ***strategic*** support and 15% in need of ***intensive*** support according to posttest results. In addition, 65% of students in the treatment group were classified as being at ***benchmark*** following the posttest.

For the control group, 73% of participants were considered in need of ***strategic*** support following the pretest, while 27% were considered in need of ***intensive*** support. According to posttest results, 59% of students were still in need of ***strategic*** support and 23% were still in need of ***intensive*** support. Eighteen percent of students in the control group were considered at ***benchmark*** following the posttest, indicating no need for supplemental instruction.

Instructional status recommendations at pretest and posttest were examined for each of the groups. Changes in status recommendations between pretest and posttest were used to place participants into one of three categories: students who moved up one status level (i.e., intensive to strategic, strategic to benchmark), students whose recommendation status did not change, and students who moved down one status level (i.e., strategic to intensive).

The results of the status recommendation change analysis are presented in Table 6. To assess whether there was a difference in status level change between the treatment and control groups, status change and group were entered into a Pearson chi-squared analysis. Of the students in the control group, 1 student (5.0%) went down one status level, while 15 (75%) stayed at the same level between pretest and posttest. Four students (20%) in the control group went up one level. No students in the treatment group went down a status level and only 6 (27.3%) remained at the same level between pretest and posttest. However, 16 students (72.7%) increased their status level by one. Chi-square analysis of this distribution indicated a significant difference ($x^2[df = 2] = 11.99$, $p =$

.002) between the groups. The students in the treatment group were significantly more likely to move up a status recommendation level following task training.

Table 6. *Instructional status recommendation change by group*

	Decrease 1 Level		No Change		Increase 1 Level	
	N	%	N	%	N	%
Treatment	0	0	6	27.3	16	72.7
Control	1	5	15	75	4	20
Total	1	2.4	21	50	20	47.6

Effects of Task Training on Individual Subtest Performance

Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of initial sounds isolated on the DIBELS Initial Sound Fluency subtest?

Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of phonemes identified on the DIBELS Phoneme Segmentation Fluency subtest?

Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of letter sounds identified on the DIBELS Nonsense Word Fluency subtest?

A one-way Hotelling's T^2, the equivalent of a MANOVA for two groups, was computed using the SPSS general linear model. The procedure simultaneously compared the independent variable (group: treatment and control) across the three dependent variables including the change score between pretest and posttest for the number of initial sounds correctly isolated, number of phonemes correctly identified, and number of letter sounds correctly identified. The multivariate approach statistically controlled for possible Type I error. The means and standard deviations for subtest performance by group are shown in Table 7. In addition, the mean change scores, change standard deviations, and effect sizes are also shown.

Table 7. *Descriptive statistics for each measure and change scores*

	Pretest		Posttest				
	Mean	SD	Mean	SD	Change Mean	Change SD	Effect Size
ISF							
Treatment	14.23	6.56	22.95	9.90	8.73	6.43	*d*=1.38
Control	15.90	8.98	14.05	6.81	-0.75	6.87	
PSF							
Treatment	9.64	9.45	20.27	9.84	10.64	10.84	*d*=2.60
Control	8.00	9.54	7.85	9.59	-0.15	4.15	
NWF							
Treatment	6.50	5.83	17.77	13.51	11.27	11.74	*d*=1.65
Control	7.05	7.50	10.20	10.42	3.75	4.56	
no task training							
LNF							
Treatment	22.73	15.50	28.55	15.05	5.82	10.93	*d*=0.33
Control	18.95	14.41	26.80	15.97	7.85	6.12	

Prior to the analyses, data were screened for normality, outliers, and missing data. Results indicated there were no missing values. The absolute value of skewness for the treatment group performance on the Phoneme Segmentation Fluency and Nonsense Word Fluency subtests was a little greater than 1 at 1.23 and 1.03, respectively. Also, skewness for the control group on the Initial Sound Fluency subtest was above an absolute value of 1 at -1.37. Kurtosis indices for the control group on the Initial Sound Fluency and Phoneme Segmentation Fluency subtests were 2.90 and 2.15, respectively. These values were also elevated on the same two measures for the treatment group with values of 1.04 and 1.06. These departures from normality were not considered serious as a result of group sample sizes being fairly equivalent.

After examining the descriptive statistics, along with a visual scan of boxplots, 11 univariate outliers were identified. It was decided to include the outliers in data analysis because differences of this nature would be expected due to the research design. To check for multivariate normality, the SPSS Regressions procedure was used to calculate Mahalanobis distance. The intent was to determine if score patterns across all dependent variables were similar and to detect any potential outliers in the full data set. It was determined that there were no multivariate outliers in the data set.

The assumption of equality of covariance matrices was not tenable (Box's M=34.01, F=5.20, p<.001), but because of fairly equivalent sample sizes the test statistic is considered robust. There was a statistically significant difference between the treatment and control group on the amount of change between pretest and posttest scores (Hotelling's T^2=30.68, F=9.72, p<.001), with a moderate effect size (η^2=.434). This result indicates that students in the group that participated in task training outperformed the control group on the combination of the three dependent variables.

One way ANOVAs were performed to examine differences between the groups for each dependent variable. Table 8 displays the results of the ANOVAs. A significance level of .05 was used for all statistical analyses. There was a significant difference between the treatment and control groups on the number of initial letter sounds isolated on the ISF posttest ($F(1, 40) = 11.302$, $p = .002$) and there was also a significant difference between the two groups on the number of phonemes correctly identified on the PSF posttest ($F(1, 40) = 17.095$, $p = .000$). On the NWF posttest, which measured the number of letter sounds identified, differences between the two groups approached significance ($F(1, 40) = 4.075$, $p = .050$).

Table 8. *ANOVA results of posttest scores by group*

	SS	*df*	MS	*F*	*p*
Initial Sound Fluency					
Between Groups	830.67	1	830.67	11.302	.002
Within Groups	2939.91	40	73.498		
Total	3770.57	41			
Phoneme Segmentation Fluency					
Between Groups	1616.73	1	1616.73	17.095	.000
Within Groups	3782.91	40	94.573		
Total	5399.64	41			
Nonsense Word Fluency					
Between Groups	600.77	1	600.77	4.075	.050
Within Groups	5897.06	40	147.427		
Total	6497.83	41			

Social Validity

The last research question addressed the acceptability and feasibility of the task training procedures. At the conclusion of the study, the 6 kindergarten teachers at the participating school observed a video-recorded demonstration of each of the task trainings. Following the video, teachers were given a questionnaire that included seven open-ended questions related to the task training procedures.

Question 1: Do you think task training with the DIBELS ISF subtest would help your students better understand what was being asked of them during the midyear benchmark screening? Please explain.

All of the teachers indicated that the task training for ISF would help their students better understand the task. The teachers further stated that the task training procedures help the students become familiar with the task directions, especially at the beginning of the school year since that is when the kindergarten students take the ISF subtest for the first time. One teacher further explained that the task training would allow the test to “better measure their ability.”

Question 2: Do you think task training with the DIBELS PSF subtest would help your students better understand what was being asked of them during the midyear benchmark screening? Please explain.

Five of the six teachers indicated that they thought the PSF task training would benefit their students. Three of the teachers indicated that they thought the strategy of tapping the table while segmenting the phonemes gave the students something “visual” and “concrete” to use during the assessment. One teacher also stated that this task training will help the students understand the directions of the task which would allow the teacher to “better target teach” that student. In contrast, one teacher indicated that tapping the table was a strategy that she already taught in her classroom, but she did not think the use of this strategy would benefit the student during the assessment.

Question 3: Do you think task training with the DIBELS NWF subtest would help your students better understand what was being asked of them during the midyear benchmark screening? Please explain.

All of the teachers made positive remarks about the NWF task training procedures. One teacher indicated that the procedures were beneficial to students because of the concentration on individual sounds instead of “reading” a word. She went on to state that the additional rows of make-believe words were also beneficial because it gave practice with the directions of the task. Another teacher indicated that the procedures focused on reading each sound and in kindergarten when students try to read the whole word they usually waste time because this is a difficult skill for them. Two teachers specifically mentioned that the focus on “short” vowel sounds during task training was beneficial.

Question 4: Do you think that task training procedures are needed to help students understand expectations for the DIBELS LNF subtest? Please explain.

Three out of six teachers indicated that task training for LNF may be beneficial for students. One of these teachers indicated that the directions for the subtest are not confusing to students, but students may need help in understanding some of the task expectations including sliding their finger under each letter, automatically moving their finger to the next row, and highlighting that they should be saying letters and not numbers. One of the three teachers who thought that task training for LNF was not needed indicated that if students know the letter names they do well on the task and are not confused by task expectations.

Question 5: Do you think any resulting improvement in students' scores on the DIBELS subtests should affect their placement in appropriate reading instruction groups? Please explain.

Four of the teachers thought that the scores after participation in task training should be used to group students for intervention. One of these teachers further explained that assessing students after their participation in task training would be a "more accurate reflection of their ability rather than their understanding of the directions." One teacher indicated that since the study was conducted in the middle of the year that it would be difficult to determine whether the improvement was due to task training and that it would be better if the task training was done at the beginning of the year. One teacher chose not to answer this question.

Question 6: Would you be able to use the task training procedures you saw on the video with your kindergarten students? (a) If so, how would you incorporate the training? (b) If not, who could implement the task training?

Five of the six teachers reported that they would use the task training procedures with their students and one teacher did not answer this question. Two teachers indicated that they could use the procedures during whole group instruction and two other teachers thought they could incorporate the trainings during small group instruction. One teacher reported that task training could be incorporated into whole group or small group instruction.

Question 7: With what type of student would you be most likely to use the task training?

Two teachers reported that all students could benefit from participation in task training procedures. Three teachers indicated that task training would be most beneficial for “low” performing students or those who were “struggling” and needed remediation. One teacher specified that task training procedures would be beneficial for students who have difficulty attending and following directions or those who are “slow” to learn directions.

CHAPTER 5: DISCUSSION

This study investigated the effects of task training, targeting three early literacy measures, in order to differentiate the need for supplemental instruction from task misunderstanding for students in kindergarten who were identified as at risk for future reading difficulties. Combined performance on four DIBELS subtests leads to an instructional status recommendation for each student. These levels include (a) ***benchmark – at grade level***, (b) ***strategic – additional intervention***, and (c) ***intensive – needs substantial intervention***. In this study, students who were identified as either at the ***strategic*** or ***intensive*** levels and who met other inclusion criteria described earlier were included in the study. These students were randomly assigned to either the treatment or control group and students in the treatment group participated in task training sessions for those subtests on which their performance resulted in anything except ***established*** or ***low risk*** depending on the specific subtest. Following task training, all participants (i.e., treatment group, control group) were administered the posttest, which was composed of alternate forms of the four DIBELS subtests administered as the pretest.

The following sections discuss the results of the analyses in terms of implications for practice, results of the social validity questionnaire, limitations of the research, and suggestions for future research.

Effects of Task Training on Dependent Variables

Will there be a significant difference between students who receive task training and students who do not receive training on the DIBELS instructional status recommendation for supplemental instruction?

Results of the current study indicate that it may be possible to minimize the number of false positives identified by the DIBELS while maintaining an accurate rate of true positives through task training. Descriptive statistics were used to examine the percentage of students categorized in each instructional recommendation group (i.e., ***intensive***, ***strategic***, ***benchmark***) based on their pretest and posttest performance. Sixty-five percent of students in the treatment group moved from either the ***intensive*** or ***strategic*** level to the ***benchmark*** level following participation in task training sessions. Whereas only 18% of students in the control group made this same change in level.

These findings suggest that, with the participants in this study, DIBELS may have over-identified students as being at risk. This corresponds to findings by Hintze et al. (2003) and Nelson (2008) who examined the classification validity for the DIBELS in separate studies. Findings of the Hintze and Nelson studies suggest high sensitivity rates which ensure identification of a high percentage of true positives. Unfortunately, in order to get such high sensitivity rates, the false positive rate is also high. High false positive rates have the potential to produce some negative consequences depending on the type of decisions being made based on the results of the assessment. Some of these negative consequences include wasting instructional resources (Bishop, 2003; Jenkins, 2003; O'Connor & Jenkins, 1999; Speece, 2005), dilution of instructional services for students truly in need of intensive, explicit, and systematic instruction, and producing unnecessary parent, teacher, and/or student anxiety (Swets et al., 2000).

Another important implication is related to the methods of dynamic assessment and precorrection, discussed in detail earlier, and the potential for these strategies to assist educators in minimizing the identification of false positives. Research on dynamic assessment (Fuchs et al., 2007; O'Connor & Jenkins, 1999) suggests its utility in reducing false positives. Supplementing early literacy screening tools with dynamic assessment procedures may help to reduce false positive rates by providing students with extra knowledge and experience they are lacking at such a young age (Catts et al., 2009). However, dynamic assessment appears to be too time consuming and individualized to effectively generalize to large groups of students assessed with universal screening measures.

Precorrection procedures may also help in the identification of false positives by helping students avoid making common errors while participating in instruction (Miao et al., 2002; O'Donnell et al, 2003). Precorrection procedures usually include a list of reminders that teachers go through with students prior to an instructional lesson or assessment. Precorrection does not focus on task directions or expectations, but rather on reminding students about specific skills that they will be expected to demonstrate (e.g., modeling correct sounds prior to lesson, reminder not to stop between sounds when blending sounds in words). In the current study, task training procedures were conducted with groups of 2 or 3 students and session lengths were considerably shorter in comparison to dynamic assessment procedures reported in previous studies (Fuchs et al., 2007; O'Connor & Jenkins, 1999). Mean task training sessions ranged from 7 min to 12 min. In addition, task training protocols were developed based on subtest expectations and common errors made by students, but were not individualized for each student. This further increases the efficiency of the task training procedures.

Following task training, students in the treatment group were significantly more likely to move up a status recommendation level (e.g., ***strategic*** to ***benchmark***) than those students in the control group. Students, who were considered in need of supplemental instruction based on their pretest performance, but moved to the benchmark level following task training, may have performed poorly at first because they did not understand the task demands. Those students in the treatment group whose status recommendation level did not change following task training are most likely in need of intervention. Being able to efficiently make this differentiation, between students truly in need of supplemental instruction and students who performed poorly because they did not understand the task, is important because it allows educators to make better informed instructional decisions about their students. For example, the reduction in the false positive rate following task training during the universal screening process may strengthen an RTI model and lead to more appropriate instructional decision making.

In addition to this benefit, reducing the number of falsely identified students may lead to more appropriate allocation of financial resources for schools, including school personnel and materials. Task training is a low cost intervention (brief, limited personnel, limited materials) that has the potential to offer substantial benefits to schools and students. If students are placed in appropriate instructional groups, then they may make greater academic gains because instruction will be focused on the skills they need to learn and not skills they have already acquired.

Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of initial sounds isolated on the DIBELS Initial Sound Fluency subtest?

Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of phonemes identified on the DIBELS Phoneme Segmentation Fluency subtest?

Will there be a significant difference between students who receive task training and students who do not receive training on the correct number of letter sounds identified on the DIBELS Nonsense Word Fluency subtest?

As noted earlier, no other research on task training to reduce false positives related to the instructional level recommendations or with particular subtests could be located except for Mackiewicz et al. (2010). That study found a significant difference between the treatment and control groups on the posttest measure, indicating that students who participated in task training for PSF outperformed students in the control group. The current study included PSF task training procedures as well as two additional subtests (i.e., ISF, NWF) and results were similar to what was found during the Mackiewicz et al. investigation following analysis using a two-way ANOVA [$F(1, 43) = 5.21$, $p = .027$, $\eta^2 = .108$].

The ISF test was the only test that kindergarten students would be familiar with in a mid-year testing. While some of the participating students had had benchmark testing in the fall with this measure and then experienced regular progress monitoring using alternate forms, results showed a significant difference between the two groups when change between pretest and posttest was examined. Specifically, the treatment group outperformed the control group on this measure. This suggests that the students in the study, even with prior exposure to this particular subtest, may not have fully understood the task directions when they had taken the test previously. Task training appears to have effectively taught the students in the treatment group the task demands and expectations.

NWF showed less robust results. Perhaps this was because assessment results for this subtest depend on students' knowledge of letter-sound correspondence, which is a higher level skill when compared to phonemic awareness activities. Also, in addition to data being examined for statistical significance, practical significance of the results was also considered. Task training for the ISF and PSF subtests resulted in statistically significant differences between the treatment and control groups when change from pretest to posttest was examined. Even though there was not a statistically significant difference between the groups on the NWF posttest, the results hold practical significance for educators. Following task training, 58.8% of students in the treatment group who scored within the ***at risk*** or ***some risk*** levels on the NWF pretest moved to the ***low risk*** level based on posttest performance, indicating the possibility that these students were not really in need of supplemental instruction for this particular skill. For students in the control group, only 27.7% moved from the ***at risk*** or ***some risk*** levels to the ***low risk*** level when pretest and posttest performance were compared.

A comparison of task training mean times revealed that task training for ISF was the quickest to teach and task training for NWF took the most time. Reasons for these differences may be that the ISF subtest has fewer task demands when compared to NWF. For the ISF subtest, students have to point to a picture or identify a word that starts with a specific sound said by the examiner. It also requires students to say, or isolate, the initial sound of a word said by the examiner. Along with exposure to examiner directions, the PSF task training focuses on one skill, segmenting words into individual phonemes. In contrast, for NWF, there are several more expectations that need to be included in that particular task training protocol. First, students must understand that they are to say the letter sounds and not letter names. Also, students are to say individual sounds in make-believe words, so they must understand that they are not reading real words. They must also say the short vowel sound anytime a vowel is encountered on the subtest and another aspect of the training includes the notion that students may skip a letter if they do not know the sound. In addition, unlike the other two subtests, NWF requires the student to independently and efficiently respond to written words from left to right rather than respond to examiner-controlled auditory prompts. Finally, the NWF task training protocol requires much more individual practice than the task trainings for the other two subtests.

Another consideration is that more errors were made by students during the NWF task training when compared to the other trainings. If students cannot identify the letter sounds during the training, an increased number of error corrections is needed, which increases the amount of time required to complete the task training.

While the Mackiewicz et al. (2010) study looked only at PSF, the present study considered the combined scores across all of the early literacy measures administered at the kindergarten midyear benchmark screening. This expansion of the investigation is important because, in most schools that use DIBELS, a combination of all subtests is used to make instructional decisions. Investigating the effects of task training on a combination of the three subtests was needed in order to differentiate the need for supplemental instruction from task misunderstanding for kindergarten students. Including all three dependent variables in the current investigation has the potential for increased application for practitioners when compared to the PSF-only study.

Discussion of Social Validity Data

In general, a group of kindergarten teachers, experienced in the administration of the DIBELS subtests, indicated through a social validity questionnaire that participation in the three task training protocols would better help their students understand the directions for each subtest. All but one response indicated that student performance following participation in task training sessions should be used to determine instructional needs because assessment following participation is an accurate measure of student ability and not a measure of their students' understanding of task directions. In addition, most of the teachers reported that task training is something that they could do with their students prior to DIBELS administration.

Since teachers perceived the procedures as something they would be able to implement themselves, it appears that task training may have more practical application than dynamic assessment or precorrection procedures when used in conjunction with a screening measure. The teachers also reported that the task training procedures were important because they help to ensure that their students understand what they are being asked to do so students can demonstrate their knowledge, or what they actually know. Teachers also indicated that they could administer the task training procedures themselves, prior to screening, with all of their students. The demonstrated brevity of the task training procedures indicate that teachers can practically administer the protocols to students without sacrificing large amounts of valuable instructional time.

Limitations

Several limitations in this study are important to discuss. First, ideally all task training would occur prior to the first administration of a particular subtest. However, the ISF subtest is first administered in the fall of kindergarten. At the participating school, ISF was administered in September and some students' progress was monitored bi-weekly through administration of ISF progress monitoring probes. However, as discussed earlier, a significant difference was found between the treatment and control group on the amount of change between the ISF pretest and posttest. This limits the current investigation because students did not have prior exposure to the other two subtests and when comparing the effects of each of the task training protocols to determine overall effectiveness, students should have a similar amount of exposure to each of the subtests.

Second, a task training protocol targeting the LNF subtest was not developed for this study. Inclusion of this task training in the current study may have potentially changed the instructional status recommendation difference between pretest and posttest. In addition, half of the teachers who completed the social validity questionnaire indicated that task training targeting the LNF subtest may be useful for several reasons. During LNF administration, students often say the letter names in the first row and stop. The examiner has to remind the students to continue identifying the letters in the next row. This often happens after each row, which wastes time. Since the measure is administered for 1 min, these pauses have the potential to lower a student's score, which may lead to their being misidentified as being at ***some risk*** on this subtest. Training students to move to the next row prior to screening administration may improve the scores of students who are not truly at risk, but performed poorly because they did not understand the task demands.

Teachers also indicated that some students do not understand the importance of moving quickly through the task and their scores are lowered because they may become distracted. Including a task training component that focuses on the timing aspect of the LNF and NWF measures may help students understand the importance of staying focused on the task which, in turn, may lead to higher scores on these subtests.

Third, the task training protocols focused specifically on one early literacy screening measure, the DIBELS. The results of this study may not translate to other early literacy screening measures. For example, the results may not generalize to such measures as AIMSWEB®, a web-based benchmark and progress monitoring system. In addition, since this study was conducted with students in kindergarten, results should not be generalized to other grades. For example, effects of task training may not be the same for students in first grade during PSF and NWF administration.

Fourth, the generalizability factor for this study is low and caution should be used due to the uniqueness of this population. This study was limited to a relatively small number of students from one elementary school. Results can only be generalized to populations similar to the participants in this study, which were all members of minority ethnicity groups (90.4% African American, 4.8% Asian, and 4.8% Hispanic). In addition, none of the students in the study received services for children with disabilities or services for children with limited English proficiency (LEP). As a result, generalizability of the current study's results is limited to students with similar characteristics. A researcher replicating this study with another population of students may obtain different results. While generalizability is not reliable, the results from this study provide some information on evaluating the effectiveness of task training procedures.

Recommendations for Future Research

Conclusions should be made with some caution due to the described limitations. Future research is needed to offset some of these limitations and to extend the research questions addressed in this study. First, the ISF task training should be conducted with kindergarteners prior to the first administration at the beginning of the school year. In the current study, ISF task training was provided after most of the students in kindergarten had already participated in at least one administration of the subtest. Having students participate in task training prior to students' first exposure to the subtest would provide more information about the effectiveness of this particular task training protocol because students would have no prior experience with the task.

Second, during the current study a task training protocol for LNF was not evaluated. This decision was made because the task of naming letters on the LNF subtest is simpler and the directions are clearer than for the other three subtests. The most likely error, saying the letter sounds rather than the letter names, is addressed during the administration procedures. However, a task training protocol focused on the LNF subtest should be developed and evaluated because some of the teacher participants indicated through a social validity questionnaire that students may benefit from task training for this subtest. Task training for LNF should include instruction on moving to the next row each time students come to the end of a row. Also, students should also be taught the importance of moving briskly through the task.

Third, since the task training protocols used in this study were developed specifically for use with the DIBELS subtests, task training protocols could be developed for other screening tools, including additional early literacy measures in order to determine if task training with these measures can help minimize the identification of false positives. These same procedures could be used with other screening measures that are parallel tests that use the same directions (i.e., AIMSweb® Phoneme Segmentation Fluency and Nonsense Word Fluency) or modified to fit changes in directions or slight differences in tasks. Also, since only one other study has been conducted investigating the effects of task training, additional research on the use of this strategy for other purposes should be considered. For example, task training protocols could be developed for other measures including tools for screening math skills.

Fourth, a duplication of this study with a larger and more varied sample size is recommended to validate the findings of the current study. Specifically, task training with DIBELS should be conducted with additional populations including students whose native language is not English. The task training protocols may need to be modified and then evaluated to determine the effectiveness for English language learners (ELL).

Fifth, future research should extend the current study by continuing to administer progress monitoring probes to students who moved from being at the ***strategic*** instructional level based on pretest scores to the ***benchmark*** level following posttest. Continuing to monitor progress would determine if these students were making adequate progress while receiving core instruction and no supplemental intervention. Extending the research in this way would help determine if the students who moved from the ***strategic*** level to the ***benchmark*** level were, in fact, false positives.

Conclusion

Accurate assessment of students' early literacy skills, especially skills that are predictive of future reading achievement, is necessary in order to make appropriate instructional decisions and provide supplemental instruction that matches student needs. Due to purposeful overidentification and floor effects, many screening measures for young children result in the identification of a high number of false positives. Task training is an abbreviated combination of dynamic assessment and precorrection that provides efficient instruction focused on helping students understand task demands. The current study used a group experimental design to determine the effects of task training, targeting three early literacy measures, in order to differentiate the need for supplemental instruction from task misunderstanding for students in kindergarten.

Task training appears to be an efficient and effective protocol that can be used to ensure correct placement of students, reducing the number of students misidentified as needing supplemental instruction. Students in the treatment group were more likely than students in the control group to move up a status recommendation level (e.g., ***strategic*** to ***benchmark***) following participation in task training. In addition, according to an analysis using a combination of scores from all three subtests, students who participated in task training outperformed the control group on the posttest.

Based on the results of this study, task training appears to have the potential to reduce the number of false positives identified when used in conjunction with the DIBELS measures at midyear kindergarten. Reducing the number of falsely identified students may have a positive impact on several aspects in a school setting and may benefit individual students as well.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Adams, M. J., Foorman, B. R., Lundberg, I., & Beeler, T. (1998). The elusive phoneme: Why phonemic awareness is so important and how to help children develop it. *American Educator, 22,* 18-29.

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 24*(1), 49-66.

Bay Area Reading Task Force (1997). *A reading-writing-language source book for the primary grades.* San Francisco, CA: University School Support for Educational Reform.

Beaver, J. (1997). *Developmental reading assessment, K-3.* Glenview, IL: Celebration Press.

Big Ideas in Beginning Reading (2009). *Alphabetic principle in beginning reading. What is the alphabetic principle? Definitions.* Retrieved June 5, 2009 from: http://reading.uoregon.edu/au/au_what_2.php.

Bishop, A. G. (2003). Prediction of first-grade reading achievment: A comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly, 26*, 189-200.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76*, 610-622.

Blachman, B. A., Ball, E. W., Black, R. S. & Tangel, D. M. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms. *Reading and Writing: An Interdisciplinary Journal, 6*, 1-18.

Blachman, B. A., Tangel, D. M., Ball, E. W., Black, R. S., & McGraw, C. K. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. *Reading and Writing: An Interdisciplinary Journal, 11*. 239-273.

Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first grade reading instruction. *Reading Research Quarterly, 2*, 5-142.

Brown-Chidsey, R., & Steege, M. W. (2005). Response to Intervention: Principles and strategies for effective practice. The Guildford Press: New York, NY.

Budoff, M. (1967). Learning potential: A supplementary procedure for assessing the ability to reason. *Seminar in Psychiatry, 1*, 278-290.

Budoff, M. (1987a). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic Testing* (pp. 173-195). New York: Guilford Press.

Budoff, M. (1987b). The validity of learning potential. In C. S. Lidz (Ed.), *Dynamic Testing* (pp. 52-81). New York: Guilford Press.

Budoff, M., & Friedman, M. (1964). "Learning potential" as an assessment approach to the adolescent mentally retarded. *Journal of Consulting Psychology, 28*, 434-439.

Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy indicators from the middle of kindergarten to second grade. *The Journal of Special Education, 42*. 209-226.

Burns, M. K., & VanDerHeyden, A.M (2006). Using response to intervention to assess learning disabilities. *Assessment for Effective Intervention, 32*(1), 3-5.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education, 41*, 254-270.

Campbell, C., & Carlson, J. S. (1995). The dynamic assessment of mental abilities. In J. S. Carlson (Ed.), *Advances in cognition and educational practice: Vol. 3. European contributions to dynamic assessment.* London: JAI Press.

Campione, J. C. (1989). Assisted testing: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities, 22*, 151-165.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic testing with school achievement. In C. S. Lidz (Ed.), *Dynamic Testing* (pp. 82-115). New York: Guilford Press.

Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in the flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9*, 297-315.

Carlson, J. S., & Wiedl, K. H. (1978). Use of testing-the-limits procedures in the testing of intellectual capabilities in children with learning difficulties. *American Journal of Mental Deficiency, 11*, 559-564.

Carlson, J. S., & Wiedl, K. H. (1979). Toward a differential testing approach: Testing-the-limits employing the Raven matrices. *Intelligence, 3*, 323-344.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*, 163-176.

Center for the Improvement of Early Reading Achievement. (2001). *Put Reading First: The research building blocks for teaching children to read*. Washington, DC: Partnership for Reading.

Chinn, P. C., & Hughes, S. (1987). Representation of minority students in special education classes. *Remedial and Special Education, 8*, 41-46.

Clay, M. M. (1985). *The early detection of reading difficulties.* Exeter, NH: Heinemann.

Colvin, G., Sugai, G., Good, R. H., & Lee, Y. Y. (1997). Using active supervision and precorrection to improve transition behaviors in an elementary school. *School Psychology Quarterly, 12*, 344-363.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*, 394-409.

Compton, D. L., Fuchs, D., Fuchs, L. S., Elleman, A. M., & Gilbert, J. K. (2008). Tracking children who fly below the radar: Latent transtion modeling of students with late-emerging reading disability. *Learning and Individual Differences, 18*, 329-337.

Cormier, P., Carlson, J. S., & Das, J. P. (1990). Planning ability and cognitive performance: The compensatory effects of a dynamic assessment approach. *Learning and Individual Differences, 2*, 437-449.

Coyne, M. D., & Harn, B. A. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools*, *43*, 33-43.

Coyne, M. D., Kame’enui, E. J., & Simmons, D. C. (2001). Prevention and intervention in beginning reading: Two complex systems. *Learning Disabilities Research & Practice*, *16*(2), 62-73.

Daly, E. J., III, Wright, J. A., Kelly, S. Q., & Martens, B. K. (1997). Measures of early academic skills: Reliability and validity with a first grade sample. *School Psychology Quarterly, 12*, 268-280.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507-524.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18,* 116-125.

Ehri, L., Nines, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250-287.

Elliott, J. (2003). Dynamic assessment in educational settings: Realizing potential. *Educational Review, 55,* 15-32.

Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills-Modified. *School Psychology Review, 30*, 33-49.

Embertson, S. E. (1992). Measuring and validating cognitive modifiability as an ability: A study in the spatial domain. *Journal of Educational Measurement, 29,* 25-50.

Fagundes, D. D., Haynes, W. O., Haak, N. J., & Moran, M. J. (1998). Task variability effects on the language test performance of southern lower socioeconomic class African American and Caucasian five-year-olds. *Language, Speech, and Hearing Services in Schools, 29,* 148-157.

Farkas, G. (2003). Racial disparities and discrimination in education: What do we know, how do we know it, and what do we need to know? *Teachers College Record, 105,* 1119-1146.

Felton, R. H. (1992). Early identification of children at risk for reading disabilities. *Topics in Early Childhood Special Education, 12*, 212-229.

Feuerstein, R., Rand, Y., & Rynders, J. (1988). *Don't accept me as I am.* New York: Plenum.

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54,* 304-331.

Fletcher, J. M., Foorman, B. R., Boudousquie, A., Barnes, M. A., Schatschneider, C., & Francis, D. J. (2002). Assessment of reading and learning disabilities: A research-based intervention-oriented approach. *Journal of School Psychology, 40,* 27-63.

Fletcher, J. M., Satz, P., & Morris, R. (1984). The Florida Longitudinal Project: A review. In S. A. Mednick, M. Harway, & K. M. Finnello (Eds.), *Handbook of longitudinal research* (Vol. 1, pp. 280-304). New York: Praeger.

Florida Center for Reading Research (2006). *Glossary of reading terms.* Retrieved June 5, 2009 from: http://www.fcrr.org/Curriculum/glossary/glossaryOfReading.pdf.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37-55.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16,* 203-212.

Francis, D., Shaywitz, S., Stuebing, K., Shaywitz, B., & Fletcher, J. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88,* 3-17.

Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., & Hill, L. (2007). Dynamic assessment as responsiveness to intervention. A scripted protocol to identify young at-risk readers. *Teaching Exceptional Children, 39,* 58-63.

Good, R. H., Gruba, J., & Kaminski, R. A. (2001). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 679-700). Washington, DC: National Association of School Psychologists.

Good, R. H., & Kaminski, R. A. (2002). Nonsense Word Fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. available: http://dibels.uoregon.edu/.

Good, R. H., & Kaminski, R. A. (Eds.). (2007). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: http://dibels.uoregon.edu/

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., et al. (2004). *Technical adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development.* (Tech. Rep. No. 7). Eugene: University of Oregon.

Good, R. J., Kaminski, R. A., & Smith, S. (2002). Phoneme Segmentation Fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: http://dibels.uoregon.edu/.

Good, R. H., Laimon, D., Kaminski, R. A., & Smith, S. (2002). Initial Sound Fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic Indicators of Basic Early Literacy Skills*

(6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: http://dibels.uoregon.edu/.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance of decision making continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5,* 257-288.

Good, R. H., Simmons, D., Kame'enui, E., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR: University of Oregon.

Good, R. H., Simmons, D. C., & Smith, S. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27,* 740-753.

Gough, P. B., & Tunmer, W. E. (1986). Decoding reading and reading disability. *Remedial and Special Education, 7,* 6-10.

Griffiths, A. J., VanDerHeyden, A. M., Parson, L. B., & Burns, M. K. (2006). Practical applications of response-to-intervention research, *Assessment for Effective Intervention, 32,* (1), 50-57.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*, 75-111.

Hagans, K. S. (2008). A response-to-intervention approach to decreasing early literacy differences in first graders from different socioeconomic backgrounds: Evidence for the intervention validity of the DIBELS. *Assessment for Effective Intervention, 34,* 35-42.

Hall, S., & Moats, L. (1999). *Straight talk about reading.* Chicago: Contemporary Books.

Haney, M. R., & Evans, J. G. (1999). National survey of school psychologists regarding use of dynamic assessment and other nontraditional assessment techniques. *Psychology in the Schools, 36*, 295-304.

Harcourt Educational Measurement. (2003). Stanford Achievement Test (10th ed.). San Antonio, TX.

Hart, B., & Risley, R. T. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brookes.

Haywood, H. C., Brown, A. L., & Wingenfeld, S. (1990). Dynamic approaches to psychoeducational assessment. *School Psychology Review, 19*, 411-422.

Haywood, H. C., & Tzuriel, D. (2002). Applications and challenges in dynamic assessment. *Peabody Journal of Education, 77,* 40-63.

Hintze, J. M., Ryan, A.L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review, 32*, 541-556.

Hollenbeck, A. F. (2007). From IDEA to implementation: A discussion of foundational and future responsiveness-to-intervention research. *Learning Disabilities Research & Practice, 22*, 137-146.

Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention or assessment: A meta-analysis of racial differences. *The Journal of Special Education, 37,* 67-80.

Individuals with Disabilities Education Improvement Act (IDEIA) of 2004, PL 108-446, 20 U.S.C. §§ 1400 et seq.

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students.* Paper presented at the national Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO.

Jitendra, A. K., & Kame'enui, E. J. (1993). Dynamic assessment as a compensatory assessment approach: A description and analysis. *Remedial and Special Education, 14,* 6-18.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development, 53*, 83-93.

Kame'enui, E. J., Fuchs, L. S., Francis, D. J., Good, R. S., O'Connor, R. E., Simmons, D. C., et al. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher, 35*(4), 3-11.

Kaminski, R. A., & Good, R. H., III (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, *25*, 215-227.

Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. Shinn (Ed.) *Advanced applications of curriculum-based measurement.* New York, NY: Guilford Press.

Kaminski, R. A., & Good, R. H. (2002). Letter Naming fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. available: http://dibels.uoregon.edu/.

Kar, B. C., Dash, U. N., Das, J. P., & Carlson, J. (1993). Two experiments on the dynamic assessment of planning. *Learning and Individual Differences, 5*, 13-29.

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of diagnostic performance of self-administred questionnaires: Extent of the problem, recommendations for standard presentation of results, and implications for the peer review process. *Psychological Assessment, 5*, 395-399.

Lewis, T. J., Colvin, G., & Sugai, G. (2000). The effects of pre-correction and active supervision on the recess behavior of elementary students. *Education and Treatment of Children, 23*, 109-121.

Lidz, C. S., & Pena, E. D. (1996). Dynamic assessment: The model, its relevance as a nonbiased approach, and its application to Latino American preschool children. *Language, Speech, and Hearing Services in Schools, 27*, 367-72.

Lopez, R. (1997). The practical impact of current research and issues in intelligence test interpretation and use for multicultural populations. *School Psychology Review, 26*, 249-254.

Lyon, G. R. (1998). Why reading is not a natural process. *Educational Leadership, 55*(6), 14-18.

MacDonald, G. W., & Cornwall, A. (1995). The relationship between phonological awareness and reading and spelling achievement eleven years later. *Journal of Learning Disabilities, 28*, 523-527.

Mackiewicz, S. M., Cooke, N. L., Galloway, T. W., & Helf, S. (2010). *Using phoneme segmentation task training to differentiate the need for supplemental instruction from task training.* Manuscript in preparation.

Manset-Williamson, G., St. John, E., Hu, S., & Gordon, D. (2002). Early literacy practices as predictors of reading related outcomes: Test scores, testing passing rates, retention, and special education referral. *Exceptionality, 10*, 11-28.

Mathes, P. G., & Torgesen, J. K. (1998). All children can learn to read: Critical care for the prevention of reading failure. *Peabody Journal of Education, 73,* 317-340.

Miao, Y., Darch, C., & Rabren, K. (2002). Use of precorrection strategies to enhance reading performance of students with learning and behavior problems. *Journal of Instructional Psychology, 29*(3), 162-174.

Missiuna, C., & Samuels, M. T. (1989). Dynamic assessment of preschool children with special needs: Comparison of mediation and instruction. *Remedial and Special Education, 10,* 53-62.

Moats, L. C. (1999). Teaching reading is rocket science! What expert teachers of reading should know and be able to do. Washington, DC: American Federation of Teachers.

Moore-Brown, B., Huerta, M., Uranga-Hernandez, Y., & Pena, E. D. (2006). Using dynamic assessment to evaluate children with suspected learning disabilities. *Intervention in School and Clinic, 41*, 209-217.

National Center for Education Statistics (2006). *The Condition of Education*. Washington: DC, National Academy Press.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel* (NIH Pub. No. 00-4754). Washington, DC: U.S. Government Printing Office.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Heath and Human Development.

National Research Center on Learning Disabilities. (2007). *Understanding responsiveness to intervention in learning disabilities determination.* Retrieved from http://www.nrcld.org

National Research Center on Response to Intervention. (2009). *What is RTI?* Retrieved October 15, 2009 from http://www.rti4success.org

Nelson, J. M. (2008). Beyond correlational analysis of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A classification validity study. *School Psychology Quarterly*, *23*, 542-552.

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. §§ 6301 *et seq*.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Connor, R. E., Fulmer, D., Harty, K. R., & Bell, K. M. (2005). Layers of reading intervention in kindergarten through third grade: Changes in teaching and student outcomes. *Journal of Learning Disabilities, 38*, 440-455.

O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. *Journal of Learning Disabilities, 38*, 532-538.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disability in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159-197.

O'Connor, R. E., Jenkins, J. R., & Slocum, T. A. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. *Journal of Educational Psychology, 2*, 202-217.

O'Connor, R. E., Notari-Syverson, N., & Vadasy, P. (1996). Ladders to literacy: The effects of teacher-led phonological activities for kindergarten children with and without disabilities. *Exceptional Children, 63,* 117-130.

O'Donnell, P., Weber, K. P., & McLaughlin, T. F. (2003). Improving correct and error rate and reading comprehension using key words and previewing: A case report with a language minority student. *Education and Treatment of Children, 26*, 237-254.

Olswang, L. B., & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech and Hearing Research, 39,* 414-423.

Oswald, K., Safran, S., & Johanson, G. (2005). Preventing trouble: Making schools safer places using positive behavior supports. *Education & Treatment of Children, 28*, p. 265-278.

Pepper, P. P., & Felton, R. H. (1995). Early identification of phonological deficits in kindergarten and early elementary children at risk for reading disability. *School Psychology Review, 24*, 405-414.

Reid, D. K., Hresko, W. P., & Hammill, D. D. (2001). Test of Early Reading Ability (3rd ed.). Austin, TX: Pro-Ed, Inc.

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the Dynamic Indicators of Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review, 35,* 241-355.

Satz, P., Taylor, H. G., Friel, J., & Fletcher, J. M. (1978). Some developmental and predictive precursors of reading disabilities: A six-year follow-up. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 457-507). New York: Oxford.

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first grade reading: Selected results from a longitudinal study. *Mental Retardation and Developmental Disabilities, 2,* 54-63.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 264-282.

Schatschneider, C., & Torgesen, J. (2004). Using our current understanding of dyslexia to support early identification and intervention. *Journal of Child Neurology, 16*, 759-765.

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*, 1309-1324.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). *Evidence that dyslexia may represent the lower tail of a normal distribution of reading disability.* The New England Journal of Medicine, 326, 145-150.

Simmons, D. C., Kame'enui, E. J., Coyne, M. D., & Chard, D. J. (2002). Effective strategies for teaching beginning reading. In E. Kame'enui, D. Carnine, R. C. Dixon, D. C. Simmons, & M. D. Coyne (Eds.), *Effective teaching strategies that accommodate diverse learners* (2rd ed., pp. 53–92). Columbus, OH: Merrill/Prentice Hall.

Slocum, T. A., O'Connor, R. E., & Jenkins, J. R. (1993). Transfer among phonological manipulation skills. *Journal of Educational Psychology, 85*, 618-630.

Smith, S., Simmons, D. C., & Kame'enui, E. J. (1998). Phonological awareness: Research bases. In D. C. Simmons & E. J. Kame'enui (Eds.). *What reading research tells us about children with diverse learning needs: Bases and basics.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Snow, C. E., Burns, M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy of Sciences.

Speece, D. L. (2005). Hitting the moving target known as reading development: Some thoughts on screening children for secondary interventions. *Journal of Learning Disabilities, 38*, 487-493.

Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology, 93*, 735-749.

Stage, S. A., Sheppard, J., Davidson, M. M., & Browning, M. M. (2001). Prediction of first-graders' growth in oral reading fluency using kindergarten letter fluency. *Journal of School Psychology, 39*, 225-237.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Swanson, H. L. (1995). Effects of dynamic testing on the classification of learning disabilities: The predictive and discriminate validity of the Swanson-Cognitive Processing Test (S-CPT). *Journal of Psychoeducational Assessment, 13*, 204-229.

Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research, 71*, 321-363.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1-26.

Tawney, J. W., & Gast, D. L. (1984). *Single subject research in special education.* Columbus, Ohio: Merrill.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*, 7-26.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voelher, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33-58, 78.

Torgesen, J. K., & Bryant, B. R. (2004). Test of Phonological Awareness Plus (2nd ed.). Austin, TX: Pro-Ed.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T. et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 1-15.

Tzuriel, D., & Shamir, A. (2002). The effects of peer mediation with young children (PMYC) on children’s cognitive modifiability. *British Journal of Educational Psychology, 77*, 143-165.

U. S. Department of Education, Office of Inspector General. (2003, February). *The department’s administration of selected aspects of the Reading First program: Final audit report* (ED-OIG/A03G0006). Available at http://www.ed.gov/about/offices/list/oig/auditreports/a03g0006.pdf

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*, 225-256.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75*, 165-183.

Vaughn, S., Wanzek, J., Woodruff, A. L., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. Klingner, & S. Vaughn (Eds.). *Evidence-Based Reading Practices for Response to Intervention* (pp. 29-42). Baltimore, MD: Brookes.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early identification as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601-638.

Vellutino, F. R., Scanlon, D., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade intervention. *Journal of Learning Disabilities, 39*, 157-169.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30*, 73-87.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing.* Austin, TX: PRO-ED.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*, 468-479.

Walsh, D. J., Price, G. G., & Gillingham, M. G. (1988). The critical but transitory importance of letter naming. *Reading Research Quarterly, 23*, 108-122.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541-561.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*, 415-438.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside.

Yopp, H. K. (1992). Developing phonemic awareness in young children. *Reading Teacher, 45*, 696-703.

APPENDIX A: TASK TRAINING SCRIPTS

Initial Sound
Fluency Task Training

**Each student will be given a student sheet with 4 pictures.

<u>Task: Picture Naming</u>

MODEL

My turn. I will name these pictures.
While pointing to each picture: ***This is bus, mailbox, chair, key.***

LEAD

Students and interventionist: while pointing to each picture
This is bus. Everybody together. What is this? (signal).
This is mailbox. What is this? (signal)
This is chair. What is this? (signal)
This is key. What is this? (signal)

TEST

Students only, while interventionist points to each picture**.**
Your turn. What is this? (signal)
Students say the name of each picture while interventionist points.

***Each student is given a turn to name all pictures independently.

<u>Task: Identifying Initial Sounds</u>

MODEL

My turn. (holding up picture of pig) ***I will say the name of the picture and then tell if it begins with /p/.***
Pig begins with /p/. Listen. Pig, /p/.
Next picture: flower. (holding up picture of flower)
Flower does not begin with /p/.

While pointing to each corresponding picture:
My turn again. I will say the name of the picture and then tell if it begins with /b/.
Bus begins with /b/. Listen. /b/, bus.

Listen. Mailbox begins with /m/. Listen. /m/, mailbox.

Listen. Chair begins with /ch/. Listen. /ch/, chair.

Listen. Key begins with /k/. Listen. /k/, key.

LEAD

Everybody together. What is this? (pointing to bus)
What is this? (pointing to mailbox)
What is this? (pointing to chair)
What is this? (pointing to key)

Does key begin with /k/? Yes, key begins with /k/.
Let's say /k/.
Do it with me /k/.
Point to the picture that begins with /k/. Students and interventionist point to or say key. ***Yes, key begins with /k/.***

Does mailbox begin with /m/? Yes, mailbox begins with /m/.
Let's say /m/.
Do it with me /m/.
Point to the picture that begins with /m/. Students and interventionist point to or say mailbox. ***Yes, mailbox begins with /m/.***

TEST
Students only - while interventionist points to the pictures.
Your turn. What is this? (while pointing to the picture of bus)
Does bus begin with /b/?
Yes, bus begins with /b/. Good job.

Point to the picture that begins with /k/.
Students point to key. ***Yes, key begins with /k/.***

Point to the picture that begins with /ch/.
Students point to chair. ***Yes, chair begins with /ch/.***

Point to the picture that begins with /m/.
Students point to mailbox. ***Yes, mailbox begins with /m/.***

***Each student is then given an opportunity to do one item independently.

<u>Task: Producing Initial Sounds</u>
MODEL
Show picture of snowman. ***My turn. I will say the sound that snowman begins with - /s/. Listen, /s/, snowman.***

Show picture of cap. ***My turn. I will say the sound that cap begins with - /k/. Listen, /k/, cap.***

LEAD
Bus begins with /b/.
Everybody together. What sound does bus begin with?
Students and interventionist say /b/.
Yes, bus begins with /b/.

Mailbox begins with /m/. What sound does mailbox begin with*? /m/*
Yes, farm begins with /f/.

Chair begins with /ch/. What sound does chair begin with*? /ch/*
Yes, chair begins with /ch/.

Key begins with /k/. What sound does key begin with*? /k/*
Yes, key begins with /k/.

TEST
Your turn.
What sound does bus begin with? Students say /b/. ***Yes, bus begins with /b/.***
What sound does mailbox begin with? Students say /m/. ***Yes, mailbox begins with /m/.***
What sound does chair begin with? Students say /ch/. ***Yes, chair begins with /ch/.***
What sound does key begin with? Students say /k/. ***Yes, key begins with /k/.***

***Each student is given an opportunity to do one item independently.

<u>Error Correction Procedures</u>

Repeat the model, lead, test steps for the particular task where the error was made.

Error correction procedures will be administered to the whole group even on individual turn errors.

Error correction procedures will be conducted only **<u>one</u>** time for each error.

Phoneme Segmentation Fluency
Task Training Procedures

**interventionist will NOT pause between sounds.

Model

A. Point to picture of the sun.
This is a picture of the sun. What is this? Students answer.
***I can say it the fast way, sun, or I can say each sound in the word.* /s/-/u/-/n/.**
Say the sounds while putting up one finger for each sound.

While holding up fingers:
How many sounds are in sun? Students answer.
***I will tap the table 3 times while saying the sounds in the word, one tap for each sound. The sounds in sun are* /s/-/u/-/n/.**
Say the sounds while tapping the table.

B. Point to picture of ice.
This is a picture of ice. What is this? Students answer.
***I can say it the fast way, ice, or I can say each sound in the word.* /i/-/s/.**
Say the sounds while putting up one finger for each sound.

While holding up fingers:
How many sounds are in ice? Students answer.
I will tap the table 2 times while saying the sounds in the word, one tap for each sound. The sounds in ice are /i/ - /s/.
Say the sounds while tapping the table.

Lead
A. Point to picture of book.
This is a picture of a book. What is this? Students answer.
/b/-/oo-/k/. Put up one finger for each sound.
How many sounds are in book? Students answer.
So altogether, how many times will we tap the table?
Right, one tap for each sound.
Let's all tap the table one time for each sound while saying the sounds in book.
Ready? Go… /b/-/oo-/k/.
Say the sounds with the students while tapping the table.

B. Point to picture of fan.
This is a picture of a fan. What is this? Students answer.
/f/-/a/-/n/ Put up one finger for each sound.
How many sounds are in fan? Students answer.
So altogether, how many times will we tap the table?
Right, one tap for each sound.
Let's all tap the table one time for each sound while saying the sounds in fan.
Ready? Go… /f/-/a/-/n/
Say the sounds with the students while tapping the table.

C. Remove the pictures.
Everybody get ready to tap the table one time for each sound.
Tell me the sounds in eat. Ready? Go…. /e/-/t/
Teacher and students say the sounds together while tapping.

Everybody get ready to tap the table one time for each sound.
Tell me the sounds in mom. Ready? Go…. /m/-/o/-/m/
Teacher and students say the sounds together while tapping.

Test
A. Tell me the sounds in zoo. Ready? Go…
All students say the sounds and tap.
Yes, the sounds in zoo are /z/-/oo/.

B. Tell me the sounds in sit. Ready? Go…
All students say the sounds and tap.
Yes, the sounds in sit are /s/-/i/-/t/.

C. Tell me the sounds in if. Ready? Go…
All students say the sounds and tap.
Yes, the sounds in if are /i/-/f/.

D. Tell me the sounds in fun. Ready? Go…
All students say the sounds and tap.
Yes, the sounds in fun are /f/-/u/-/n/.

Individual Test
Each student is given two words (any combination of the following).
(Child's Name), your turn to tell me the sounds in no.
Yes, the sounds in no are /n/-/o/.
(Child's Name), your turn to tell me the sounds in man.
Yes, the sounds in man are /m/-/a/-/n/.
(Child's Name), your turn to tell me the sounds in am.
Yes, the sounds in am are /a/-/m/.
(Child's Name), your turn to tell me the sounds in lip.
Yes, the sounds in lip are /l/-/i/-/p/.

Error Correction Procedures
My turn to tap and say each sound in __________. (Model saying each sound while tapping).
Get ready to tap and say the sounds in _______ ***with me. Ready? Go….***
Say the sounds with the students while tapping the table.
Your turn, tap and say the sounds in _________***. Ready? Go….***

Error correction procedures will be administered to the whole group even on individual turn errors.
Error correction procedures will be conducted only **one** time for each error.

Nonsense Word Fluency
Task Training

**Each student has a student sheet in front of them.

A. Identifying Consonant Letter Sounds

Put your finger under the number 1. Look at the first row of letters. I can say each letter's sound. Watch while I point under each letter as I say its sound. My turn.
Say each sound while pointing under the letter.

Now, you are going to point under each letter as I say its sound. Get ready, go.
Say each sound while pointing under the letter. Monitor students as they point under each letter. Provide feedback and/or error correction as needed (e.g., "good pointing").

Now it's your turn. Put your finger back under the number 1. Everybody together, say the sounds while you point under each letter. Get ready, go.
Signal. Provide feedback and/or error correction.

B. Identifying Short Vowel Letter Sounds

Now let's look at the next row. Put your finger under the number 2. We're going to use the short vowel sounds because all the words we will be reading next have short vowel sounds. This row has all vowel letters in it. I'm going to say the short vowel sounds. My turn to say each sound while you point.
Say each vowel sound while pointing under the corresponding letter.

Now it's your turn. Put your finger back under the number 2. Everybody together, say the short vowel sounds while you point under each letter. Get ready, go.
Signal. Provide feedback and/or error correction.

C. My Turn – Your Turn: Reading each word

We can put these letters together to make make-believe words. Make-believe words are not real. They are pretend words.

Put your finger under the number 3. Words in this row are make-believe, or pretend, words. Put your finger under the first word in this row. Listen as I say the sounds in this make-believe word: My turn. /t/ - /i/ - /g/. Tig is not a real word, it is a make-believe word. /t/ - /i/ - /g/
Point under each letter as you say the sound.

Keep your finger under the first word in row 3. Your turn to say the sounds in this make-believe word. Get ready, go.
Signal. Provide feedback and/or error correction.

Now put your finger under the next word. Model pointing under second word. My turn to say the sounds in this make-believe word.
Say each sound in the word while pointing.
Now it's your turn to say the sounds in this make-believe word. Get ready, go.
Provide feedback and/or error correction.

Continue **My turn – Your turn** for all words in the third row.

D. My Turn – Your Turn: Reading row of words

Put your finger under the number 4. Listen as I say the sounds in each make-believe word in this row. Follow along with your finger, pointing under each sound as I say it. My turn. Read each sound in the make-believe words.

Put your finger back under the number 4. Now it's your turn to say the sounds in all of the words in this row. First word, everybody, get ready, go.
Next word. Go. Say "next word. go" before each word in the row.
Provide feedback and/or error correction.

E. Individual Turns

When you are reading the make-believe words, you can skip a letter if you don't know the sound. Now I'm going to call on just 1 student to read some more make-believe words.

Everybody, put your finger under the number 5. Here are some more make-believe words. Start here **(point to the first word)** ***and go across the page*** **(point across the page)*****. When I say "begin," read the words the best you can. Point to each letter and tell me the sound or read the whole word. Read the words the best you can. (Student's name), put your finger on the first word. Ready, begin.*** Student says each sound in the row 5 words. As needed, remind students they can skip letters they don't know while providing feedback and/or error correction.

Everybody, put your finger under the number 6. Here are some more make-believe words. Start here **(point to the first word)** ***and go across the page*** **(point across the page)*****. When I say "begin," read the words the best you can. Point to each letter and tell me the sound or read the whole word. Read the words the best you can. (Student's name), put your finger on the first word. Ready, begin.*** Student says each sound in the row 6 words. As needed, remind students they can skip letters they don't know while providing feedback and/or error correction.

Everybody, put your finger under the number 7. Here are some more make-believe words. Start here **(point to the first word)** ***and go across the page*** **(point across the page)*****. When I say "begin," read the words the best you can. Point to each letter and tell me the sound or read the whole word. Read the words the best you can. (Student's name), put your finger on the first word. Ready, begin.*** Student says each sound in the row 7 words. As needed, remind students they can skip letters they don't know while providing feedback and/or error correction.

Error Correction Procedures

For rows 1 – 4, immediately do **My turn – Together – Your turn** when an error is made. Error correction procedures will be administered to the whole group.

For individual turns – follow the DIBELS NWF scoring rules. No error correction for missed sounds; only correct procedural errors like not skipping letters they don't know, saying the long vowel sound, or substituting real words for make-believe words.

Error correction procedures will be conducted only **one** time for each error.

APPENDIX B: TASK TRAINING TREATMENT FIDELITY CHECKLISTS

Initial Sound Fluency Task Training
Treatment Fidelity Checklist

Date: ___________ Interventionist: __________________ Session Length _____________________ Checklist completed by __________________

For each step, circle YES if it occurred, NO if it did not occur, or NA if the step is not applicable.

I = Interventionists; S=Students

Phase 1: Picture Naming		
MODEL		
1. I = "My turn. I will name these pictures"	YES	NO
2. I = Says the name of each picture while pointing to it.	YES	NO
LEAD		
4. I = "This is bus. Everybody together. What is this?"	YES	NO
5. I = "This is mailbox. Everybody together. What is this?"	YES	NO
6. I = "This is chair. Everybody together. What is this?"	YES	NO
7. I = "This is key. Everybody together. What is this?"	YES	NO
TEST		
8. I = "Your turn. What is this?"	YES	NO
9. S = Students say the names of the pictures while pointing.	YES	NO
10. Each student is given an opportunity name all pictures.	YES	NO

26. I = "Does mailbox begin with /m/?" Students answer.	YES	NO
27. I = "Let's say /m/. Do it with me, /m/." Students answer.	YES	NO
28. I = "Point to the picture that begins with /m/." Students point to pic.	YES	NO
TEST		
29. I = "Your turn. What is this?" Point to picture of bus.	YES	NO
30. I = "Does bus begin with /b/?" Students answer.	YES	NO
31. I = "Point to the picture that begins with /k/." Students point.	YES	NO
32. I = "Point to the picture that begins with /ch/." Students point.	YES	NO
33. I = "Point to the picture that begins with /m/." Students point.	YES	NO
34. Each student is given an opportunity to do one item independently.	YES	NO
Phase 3: Producing Initial Sounds		
MODEL		
35. I = Holds up picture of snowman. "My turn. I will say the sound that snowman begins with - /s/. Listen, /s/, snowman."	YES	NO

Phase 2: Identifying Initial Sounds		
MODEL		
11. I = “My turn.” Holds up picture of pig. “I will say the name of the picture and then tell if it begins with /p/.	YES	NO
12. I = “Pig begins with /p/. Listen. Pig, /p/. “	YES	NO
13. I = “Next picture: flower.” Holds up picture of flower. “Flower does not begin with /p/.”	YES	NO
14. I = “My turn again. I will say the name of the picture and then tell if it begins with /b/.”	YES	NO
15. I = “Bus begins with /b/. Listen. /b/, bus.”	YES	NO
16 .I = “Listen. Mailbox begins with /m/. Listen. /m/, mailbox.”	YES	NO
17. I = “Listen. Chair begins with /ch/. Listen. /ch/, chair.”	YES	NO
18. I = “Listen. Key begins with /k/. Listen. /k/, key.”	YES	NO
LEAD		
19. I = “Everybody together. What is this?” Point to bus.	YES	NO
20. I = “What is this?” Point to mailbox.	YES	NO
21. I = “What is this?” Point to chair.	YES	NO
22. I = “What is this?” Point to key.	YES	NO
23. I = “Does key begin with /k/?” Students answer.	YES	NO
24. I = “Let’s say /k/. Do it with me, /k/.” Students answer.	YES	NO
25. I = “Point to the picture that begins with /k/.” Students point	YES	NO

36. I = Holds up picture of cap. “My turn. I will say the sound that cap begins with - /k/. Listen, /k/, cap.	YES	NO
LEAD		
37. I = “Bus begins with /b/. What sound does bus begin with?”	YES	NO
38. I = “Mailbox begins with /m/. What sound does mailbox begin with?”	YES	NO
39. I = “Chair begins with /ch/. What sound does chair begin with?”	YES	NO
40. I = “Key begins with /k/. What sound does key begin with?”	YES	NO
TEST		
41. I = “Your turn. What sound does bus begin with?” Students answer.	YES	NO
42. I = “What sound does mailbox begin with?” Students answer.	YES	NO
43. I = “What sound does chair begin with?” Students answer.	YES	NO
44. I = “What sound does key begin with?” Students answer.	YES	NO
45. Each student is given an opportunity to do one item independently.	YES	NO
ERROR CORRECTION PROCEDURES		
46. Error correction procedures are followed as needed.	YES	NO
	NA	

Phoneme Segmentation Fluency Task Training

Treatment Fidelity Checklist

Date: _________________ Interventionist: ________________________ Session Length ______________ Checklist completed by __________________

For each step, circle YES if it occurred, NO if it did not occur, or circle NA if the step is not applicable.
I = Interventionist; S=Students

Phase 1: Model		**Picture 1**		**Picture 2**	
1. I = “This is a picture of ____. What is this?” Students answer.		YES	NO	YES	NO
2. I = “I can say it the fast way, ____, or I can say each sound in the word.”		YES	NO	YES	NO
3. I = Sounds are said while putting up one finger for each sound.		YES	NO	YES	NO
4. I = While holding up fingers: “How many sounds are in ___?” Students answer.		YES	NO	YES	NO
5. I = “I will tap the table __ times while saying the sounds in the word, one tap for each sound. The sounds in _____ are _________.”		YES	NO	YES	NO
6. I = Sounds are said while tapping the table.		YES	NO	YES	NO
Phase 2: Lead with pictures		**Picture 1**		**Picture 2**	
7. I = “This is a picture of _____. What is this?” Students answer.		YES	NO	YES	NO
8. The word is segmented and one finger is held up as each sound is said aloud.		YES	NO	YES	NO
9. I = “How many sounds are in _____?” Students answer.		YES	NO	YES	NO
10. I = “So all together, how many times will we tape the table?” Students answer.		YES	NO	YES	NO
11. I = “Right one tap for each sound. Let’s all tap the table one time for each sound while saying the sounds in _______. Ready, go.”		YES	NO	YES	NO
12. I & S = Sounds are said with the students while tapping the table.		YES	NO	YES	NO

Error Correction Procedures				
13. “My turn to tap and say each sound in _____. The sounds in ____ are _____.”		YES	NO	NA
14. “Get ready to tap and say the sounds in ____ with me. Ready, go.”		YES	NO	NA
15. Sounds are said while tapping the table.		YES	NO	NA
16. “Your turn, tap and say the sounds in _______. Ready, go.”		YES	NO	NA

Phase 3: Lead without pictures	Word 1		Word 2					
17. I = “Everybody get ready to tap the table one time for each sound. Tell me the sounds in _______. Ready, Go.”	YES	NO	YES	NO				
18. Sounds are said with the students while tapping the table.	YES	NO	YES	NO				
Error Correction Procedures								
19. “My turn to tap and say each sound in _____. The sounds in ____ are _____.”	YES		NO		NA			
20. “Get ready to tap and say the sounds in ____ with me. Ready, go.”	YES		NO		NA			
21. Sounds are said while tapping the table.	YES		NO		NA			
22. “Your turn, tap and say the sounds in _______. Ready, go.”	YES		NO		NA			
Phase 4: Test	**Word 1**		**Word 2**		**Word 3**		**Word 4**	
23. I = “Tell me the sounds in _______. Ready, go.”	YES	NO	YES	NO	YES	NO	YES	NO
24. S = All students say the sounds and tap the table.	YES	NO	YES	NO	YES	NO	YES	NO
Error Correction Procedures								
25. “My turn to tap and say each sound in _____. The sounds in ____ are _____.”	YES		NO		NA			
26. “Get ready to tap and say the sounds in ____ with me. Ready, go.”	YES		NO		NA			
27. Sounds are said while tapping the table.	YES		NO		NA			
28. “Your turn, tap and say the sounds in _______. Ready, go.”	YES		NO		NA			
Phase 5: Individual Test	**Practice**							
29. Each student is given an opportunity to practice individually with two words.	YES	NO						
Error Correction Procedures								
30. “My turn to tap and say each sound in _____. The sounds in ____ are _____.”	YES		NO		NA			
31. “Get ready to tap and say the sounds in ____ with me. Ready, go.”	YES		NO		NA			
32. Sounds are said while tapping the table.	YES		NO		NA			
33. “Your turn, tap and say the sounds in _______. Ready, go.”	YES		NO		NA			

Nonsense Word Fluency Task Training

Treatment Fidelity Checklist

Date ________ Interventionist ________________ Session Length __________
Checklist completed by ________

Circle **YES** if action occurred
Circle **NO** if action did not occurred
Circle **NA** if action was not applicable

I = Interventionist
S = student(s)

A. Identifying Consonant Letter Sounds	**Circle One**		
1. I = "Put your finger under the number 1. Look at the first row of letters. I can say each letter's sound. Watch while I point under each letter as I say its sound. My turn."	**YES**	**NO**	
2. I = Each sound in the row is said aloud while pointing under the corresponding letter.	**YES**	**NO**	
3. I = "Now you are going to point under each letter as I say its sound. Get ready, go."	**YES**	**NO**	
4. I = Each sound in the row is said aloud while students point under each letter.	**YES**	**NO**	
5. I = "Now it's your turn. Put your finger back under the number 1. Everybody together, say the sounds while you point under each letter. Get ready, go.	**YES**	**NO**	
6. I = Feedback and/or error correction is provided.	**YES**	**NO**	**NA**
B. Identifying Short Vowel Letter Sounds	**Circle One**		
7. I = "Now let's look at the next row. Put your finger under the number 2. We're going to use the short vowel sounds because all the words we will be reading next have short vowel sounds. This row has all vowel letters in it. I'm going to say the short vowel sounds."	**YES**	**NO**	
8. I = "My turn to say each sound while you point."	**YES**	**NO**	
9. I = Each vowel sound is said aloud while pointing to the corresponding letter.	**YES**	**NO**	
10. I = "Now it's your turn. Put your finger back under the number 2. Everybody together, say the short vowel sounds while you point under each letter. Get ready, go."	**YES**	**NO**	
11. S = Students say each sound in row 2.	**YES**	**NO**	
12. I = Feedback and/or error correction is provided.	**YES**	**NO**	**NA**
C. My Turn – Your Turn; Reading Each Word	**Circle One**		
13. I = "We can put these letters together to make make-believe words. Make-believe words are not real. They are pretend words. Put your finger under the number 3. Words in this row are make-believe, or pretend, words. Put your finger under the first word in this row. Listen as I say the sounds in this make-believe word.	**YES**	**NO**	
14. I = "My turn. /t/ - /i/ - /g/. Tig is not a real word. It is a make-believe word. /t/ - /i/ - /g/."	**YES**	**NO**	
15. I = "Keep your finger under the first word in row 3. Your turn to say the sounds in this make-believe word. Get ready, go."	**YES**	**NO**	
16. S = Students say each sound in /tig/.	**YES**	**NO**	
17. I = Feedback and/or error correction is provided.	**YES**	**NO**	**NA**
18. I = "Now put your finger under the next word." (I puts finger on next word).	**YES**	**NO**	
19. I = "My turn to say the sounds in the make-believe word. /o/ - /l/.	**YES**	**NO**	
20. I = "Your turn to say the sounds in the make-believe word. Get ready, go."	**YES**	**NO**	
21. S = Students say each sound in /ol/.	**YES**	**NO**	
22. I = Feedback and/or error correction is provided.	**YES**	**NO**	**NA**
23. I = "Now put your finger under the next word." (I puts finger on next word).	**YES**	**NO**	
24. I = "My turn to say the sounds in the make-believe word. /d/ - /a/ - /k/."	**YES**	**NO**	
25. I = "Your turn to say the sounds in the make-believe word. Get ready, go."	**YES**	**NO**	
26. S = Students say each sound in /dak/.	**YES**	**NO**	
27. I = Feedback and/or error correction is provided.	**YES**	**NO**	**NA**
28. I = "Now put your finger under the next word." (I puts finger on next word).	**YES**	**NO**	

29. I = "My turn to say the sounds in the make-believe word. /s/ - /e/ - /p/."		YES	NO	
30. I = "Your turn to say the sounds in the make-believe word. Get ready, go."		YES	NO	
31. S = Students say each sound in /sep/.		YES	NO	
32. I = Feedback and/or error correction is provided.		YES	NO	NA
33. I = "Now put your finger under the next word." (I puts finger on next word).		YES	NO	
34. I = "My turn to say the sounds in the make-believe word. /u/ - /b/."		YES	NO	
35. I = "Your turn to say the sounds in the make-believe word. Get ready, go."		YES	NO	
36. S = Students say each sound in /ub/.		YES	NO	
30. I = Feedback and/or error correction is provided.		YES	NO	NA
D. My Turn – Your Turn: Reading Row of Words		**Circle One**		
31. I = "Put your finger under the number 4. Listen as I say the sounds in each make-believe word in this row. Follow along with your finger, pointing under each sound as I say it. My turn."		YES	NO	
32. I = Each sound in the make-believe words is said aloud while pointing to the corresponding letter.		YES	NO	
33. I = "Put your finger back under the number 4. Now it's your turn to say the sounds in all of the words in this row. First word, everybody, get ready, go."		YES	NO	
34. I = Say "next word – go" before each word in the row.		YES	NO	
35. S = Students read each word in the row.		YES	NO	
36. I = Feedback and/or error correction is provided.		YES	NO	NA
E. Individual Turns		**Circle One**		
37. I = "When you are reading the make-believe words, you can skip a letter if you don't know the sound. Now I'm going to call on just 1 student to read some more make-believe words."		YES	NO	
38. I = "Everybody, put your finger under the number 5. Here are some more make-believe words. Start here (point to first word) and go across the page."		YES	NO	
39. I = "When I say 'begin,' read the words the best you can. Point to each letter and tell me the sound or read the whole word. Read the words the best you can."		YES	NO	
40. I = "(Student's name), you can read this row. Ready, begin."		YES	NO	
41. S = Student says each sound in the row 5 words.		YES	NO	
42. I = As needed, remind student he/she can skip letters they don't know.		YES	NO	NA
43. I = Feedback and/or error correction is provided.		YES	NO	NA
44. I = "Everybody, put your finger under the number 6. Here are some more make-believe words. Start here (point to first word) and go across the page."		YES	NO	
45. I = "When I say 'begin,' read the words the best you can. Point to each letter and tell me the sound or read the whole word. Read the words the best you can."		YES	NO	
46. I = "(Student's name), you can read this row. Ready, begin."		YES	NO	
47. S = Student says each sound in the row 6 words.		YES	NO	
48. I = As needed, remind student he/she can skip letters they don't know.		YES	NO	NA
49. I = Feedback and/or error correction is provided.		YES	NO	NA
50. I = "Everybody, put your finger under the number 7. Here are some more make-believe words. Start here (point to first word) and go across the page."		YES	NO	
51. I = "When I say 'begin,' read the words the best you can. Point to each letter and tell me the sound or read the whole word. Read the words the best you can."		YES	NO	
52. I = "(Student's name), you can read this row. Ready, begin."		YES	NO	
53. S = Student says each sound in the row 7 words.		YES	NO	
54. I = As needed, remind students they can skip letters they don't know.		YES	NO	NA
55. I = Feedback and/or error correction is provided.		YES	NO	NA

APPENDIX C: SOCIAL VALIDITY QUESTIONNAIRE

1. Do you think task training with the DIBELS ISF subtest would help your students better understand what was being asked of them during the midyear benchmark screening?

2. Do you think task training with the DIBELS PSF subtest would help your students better understand what was being asked of them during the midyear benchmark screening?

3. Do you think task training with the DIBELS NWF subtest would help your students better understand what was being asked of them during the midyear benchmark screening?

4. Do you think that task training procedures are needed to help students understand expectations for the DIBELS LNF subtest?

5. Do you think any resulting improvement in students' scores on the DIBELS subtests should effect their placement in an appropriate reading instruction groups? Please explain.

6. Would you be able to use the task training procedures you saw on the video with your kindergarten students?

    a. If so, how would you incorporate the training?

    b. If not, who could implement the task training?

7. With what type of student would you be most likely to use the task training?